

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 09:44:14 ; Search time 52 Seconds  
(without alignments)  
4465.796 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 736  
Sequence: 1 MAVRALKLLTLLAVVAAS.....PPPEAQSLSGSEKPKDEL 736

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL 25: \*  
1: sp\_archaea: \*  
2: sp\_bacteria: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phase: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp\_vertebrate: \*  
14: sp\_unclassified: \*  
15: sp\_rvirus: \*  
16: sp\_bacteriap: \*  
17: sp\_archaeap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match length	ID Description
1	736	100.0	Q9HC86 homo sapien
2	433	58.8	Q96SK8 homo sapien
3	407	55.3	Q96SL5 homo sapien
4	360	48.9	Q9HC87 homo sapien
5	289	39.3	Q96BR8 homo sapien
6	281	38.2	Q96SN3 homo sapien
7	188	25.5	Q96SK3 homo sapien
8	101	13.7	Q9RIJ8
9	101	13.7	Q9CFL3
10	101	13.7	Q9CKW5
11	91	12.4	Q9QZT7
12	82	11.1	Q9QZT6
13	80	10.9	Q9BSV2
14	16	2.2	Q8C673
15	16	2.2	Q8CG71
16	13	1.8	Q9NVI2

17	13	1.8	708	4	Q8IVL5	Q8IVL5 homo sapien
18	10	1.4	412	5	Q9NDP6	Q9NDP6 ciona intes
19	8	1.1	133	8	Q36544	Q36544 scarabaeus
20	8	1.1	187	16	Q7V309	Q7V309 prochloroc
21	8	1.1	196	16	Q8A8T9	Q8A8T9 bacteroides
22	8	1.1	222	16	Q9PP39	Q9PP39 campylobact
23	8	1.1	227	4	Q33513	Q33513 homo sapien
24	8	1.1	241	8	Q8WAQ0	Q8WAQ0 agabus bino
25	8	1.1	241	8	Q8WAQ2	Q8WAQ2 agabus bigu
26	8	1.1	241	8	Q8WSH1	Q8WSH1 agabus heyd
27	8	1.1	241	8	Q8WAP9	Q8WAP9 agabus dila
28	8	1.1	241	8	Q8WSH0	Q8WSH0 agabus bigu
29	8	1.1	241	8	Q958P2	Q958P2 agabus bipu
30	8	1.1	241	8	Q8WAQ1	Q8WAQ1 agabus bigu
31	8	1.1	241	8	Q8WAQ3	Q8WAQ3 agabus gutt
32	8	1.1	241	8	Q8WAP8	Q8WAP8 agabus gutt
33	8	1.1	241	8	Q8WAP7	Q8WAP7 agabus cf.
34	8	1.1	252	8	Q8HEV5	Q8HEV5 agabus disi
35	8	1.1	257	5	O16619	O16619 caenorhabdi
36	8	1.1	274	8	Q8SA66	Q8SA66 drosophila
37	8	1.1	288	10	Q84RS4	Q84RS4 oryza sativ
38	8	1.1	309	16	Q8PM96	Q8PM96 xanthomonas
39	8	1.1	317	16	Q9XAN8	Q9XAN8 streptomyce
40	8	1.1	388	4	Q8NB01	Q8NB01 homo sapien
41	8	1.1	443	11	Q8K2B0	Q8K2B0 mus musculu
42	8	1.1	452	16	Q8A4S2	Q8A4S2 bacteroides
43	8	1.1	457	11	Q8C1I2	Q8C1I2 mus musculu
44	8	1.1	466	8	Q9MT06	Q9MT06 isemene narc
45	8	1.1	493	12	Q993S9	Q993S9 anticarsia

ALIGNMENTS

RESULT 1  
Q9HC86  
ID Q9HC86 PRELIMINARY; PRT; 736 AA.  
AC Q9HC86;  
CT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE GROS1-L protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]\_TaxID=9606;  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=20406537; PubMed=10951563;  
RA Kaul S.C., Sugihara T., Yoshida A., Nomura H., Wachwa R.;  
RT "Gros1, a potential growth suppressor on chromosome 1: its identity to  
RT basement membrane-associated proteoglycan, leprecan.";  
RL Oncogene 19:3576-3583(2000).  
DR EMBL; AF097432; AAC31019.1; --  
DR GO; GO:0016706; F:oxidoreductase activity, acting on paired d. . .; IEA.  
DR GO; GO:0019538; P:protein metabolism; IEA.  
DR InterPro; IPR005123; 2OG-Fell Oxy.  
DR InterPro; IPR000886; ER target S.  
DR InterPro; IPR006620; Pco-4\_hyd\_alph.  
DR InterPro; IPR008941; IPR-like.  
DR Pfam; PF03171; 2OG-Fell Oxy; 1.  
DR SMART; SM00702; P4HG; 1.  
DR PROSITE; PS00014; ER TARGET; 1.  
SQ SEQUENCE 736 AA; 83413 MW; 59D1205DB76ADB50 CRC64;

Query March 100.0%; Score 736; DB 4; Length 736;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 736; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAVRALKLLTLLAVVAASQAQVESEAGWGVTPTDLLFAEGTAAYARGDWPGVLSMER 60  
Db 1 MAVRALKLLTLLAVVAASQAQVESEAGWGVTPTDLLFAEGTAAYARGDWPGVLSMER 60

QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLRRAACL 120  
DB 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLRRAACL 120  
QY 121 RRLCGPAAHLSSEMELEFRKSPYNYLQVAFKINKLEKAVAAAHATFFVGNPEHMEQ 180  
DB 121 RRLCGPAAHLSSEMELEFRKSPYNYLQVAFKINKLEKAVAAAHATFFVGNPEHMEQ 180  
QY 181 QNLDDYQTMGSVKVADPKLETPHMQEPRGLVRLYSEEQPQAVPHLEALQYFVAYE 240  
DB 181 QNLDDYQTMGSVKVADPKLETPHMQEPRGLVRLYSEEQPQAVPHLEALQYFVAYE 240  
QY 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPFDEL 300  
DB 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPFDEL 300  
QY 301 PSHNYLQFAYNYIGNYTOAGECAKTYLLFPFNDVNMNQLAYYAAWLGEEHTRSIGPRE 360  
DB 301 PSHNYLQFAYNYIGNYTOAGECAKTYLLFPFNDVNMNQLAYYAAWLGEEHTRSIGPRE 360  
QY 361 SAKYRORSLEKELLFPAYDVFGIPVDPSWTPEEVI PKRLOEKQKSERETAVRISQE 420  
DB 361 SAKYRORSLEKELLFPAYDVFGIPVDPSWTPEEVI PKRLOEKQKSERETAVRISQE 420  
QY 421 IGMLMKEIETLVEEKTESLDVSRLTREGGPLYEGISLTWNSKLLNGYQVVMGVID 480  
DB 421 IGMLMKEIETLVEEKTESLDVSRLTREGGPLYEGISLTWNSKLLNGYQVVMGVID 480  
QY 481 HECQELQRLTNVAATSGDGRGQTSPTPNEKFGYGVTVFKALKLGQEGKVPLOSALHYN 540  
DB 481 HECQELQRLTNVAATSGDGRGQTSPTPNEKFGYGVTVFKALKLGQEGKVPLOSALHYN 540  
QY 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPHVNDNCILNAETLVC 600  
DB 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPHVNDNCILNAETLVC 600  
QY 601 VKEPPAYTRDYSAILYLNGDFDGNFYFTLDAKTVTAEVQOCGRVGFSSGTENPHG 660  
DB 601 VKEPPAYTRDYSAILYLNGDFDGNFYFTLDAKTVTAEVQOCGRVGFSSGTENPHG 660  
QY 661 VKAVTRGORCAIALWFTLDPHRSERDQVADDLVKMLFSPPEMDLSQEQPLDAQQPPPEP 720  
DB 661 VKAVTRGORCAIALWFTLDPHRSERDQVADDLVKMLFSPPEMDLSQEQPLDAQQPPPEP 720  
QY 721 AQESLSGSESKPKDEL 736  
DB 721 AQESLSGSESKPKDEL 736

## RESULT 2

Q96SK8 PRELIMINARY; PRT; 736 AA.  
AC Q96SK8  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein FLJ14791 (Hypothetical protein  
NT2P2005671).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_taxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,  
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,  
RA Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;  
RT "NSDO human cDNA sequencing project";  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.

RA Ota T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayashi K., Ishii S.,  
RA Saito K., Yamamoto J., Wakamatsu A., Nagai T., Nakamura Y.,  
RA Nagahari K., Sugano S., Isogai T.;  
RT "HRI human cDNA sequencing project";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK027697; BAB55305.1; -;  
DR EMBL; AK075418; BAC11608.1; -;  
DR InterPro; IPR005123; 2OG-Fell\_Oxy.  
DR InterPro; IPR000886; ER-target\_S.  
DR InterPro; IPR008941; TPR-like\_S.  
DR Pfam; PF03171; 2OG-Fell\_Oxy; 1.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 736 AA; 83393 MW; EA1909828FAB685E CRC64;  
Query Match 58.8%; Score 433; DB 4; Length 736;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MAVRALKLLTLLAVVAAASQAEESEAGWGNTPDLLFABGTAAYARGDPGVLSMER 60  
DB 1 MAVRALKLLTLLAVVAAASQAEESEAGWGNTPDLLFABGTAAYARGDPGVLSMER 60  
QY 61 ALRSPAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLRRAACL 120  
DB 61 ALRSPAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLRRAACL 120  
QY 121 RRLCGPAAHLSSEMELEFRKSPYNYLQVAFKINKLEKAVAAAHATFFVGNPEHMEQ 180  
DB 121 RRLCGPAAHLSSEMELEFRKSPYNYLQVAFKINKLEKAVAAAHATFFVGNPEHMEQ 180  
QY 181 QNLDDYQTMGSVKVADPKLETPHMQEPRGLVRLYSEEQPQAVPHLEALQYFVAYE 240  
DB 181 QNLDDYQTMGSVKVADPKLETPHMQEPRGLVRLYSEEQPQAVPHLEALQYFVAYE 240  
QY 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPFDEL 300  
DB 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPFDEL 300  
QY 301 PSHNYLQFAYNYIGNYTOAGECAKTYLLFPFNDVNMNQLAYYAAWLGEEHTRSIGPRE 360  
DB 301 PSHNYLQFAYNYIGNYTOAGECAKTYLLFPFNDVNMNQLAYYAAWLGEEHTRSIGPRE 360  
QY 361 SAKYRORSLEKELLFPAYDVFGIPVDPSWTPEEVI PKRLOEKQKSERETAVRISQE 420  
DB 361 SAKYRORSLEKELLFPAYDVFGIPVDPSWTPEEVI PKRLOEKQKSERETAVRISQE 420  
QY 421 IGMLMKEIETLVEEKTESLDVSRLTREGGPLYEGISLTWNSKLLNGYQVVMGVID 480  
DB 421 IGMLMKEIETLVEEKTESLDVSRLTREGGPLYEGISLTWNSKLLNGYQVVMGVID 480  
QY 481 HECQELQRLTNVAATSGDGRGQTSPTPNEKFGYGVTVFKALKLGQEGKVPLOSALHYN 540  
DB 481 HECQELQRLTNVAATSGDGRGQTSPTPNEKFGYGVTVFKALKLGQEGKVPLOSALHYN 540  
QY 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPHVNDNCILNAETLVC 600  
DB 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPHVNDNCILNAETLVC 600  
QY 601 VKEPPAYTRDYSAILYLNGDFDGNFYFTLDAKTVTAEVQOCGRVGFSSGTENPHG 660  
DB 601 VKEPPAYTRDYSAILYLNGDFDGNFYFTLDAKTVTAEVQOCGRVGFSSGTENPHG 660  
QY 661 VKAVTRGORCAIALWFTLDPHRSERDQVADDLVKMLFSPPEMDLSQEQPLDAQQPPPEP 720  
DB 661 VKAVTRGORCAIALWFTLDPHRSERDQVADDLVKMLFSPPEMDLSQEQPLDAQQPPPEP 720  
QY 721 AQESLSGSESKPKDEL 736  
DB 721 AQESLSGSESKPKDEL 736

```
Q96SL5
ID Q96SL5 PRELIMINARY; PRT; 736 AA.
AC Q96SL5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ14774.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027680; BAB55291.1; -.
DR GO; GO:0016706; F:oxidoreductase activity, acting on paired d. ...; IEA.
DR GO; GO:0019538; P:protein metabolism; IEA.
DR InterPro; IPR005123; 2OG-Fell Oxy.
DR InterPro; IPR000886; ER-target Oxy.
DR InterPro; IPR006620; Pro 4 hyd_alph.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; 2OG-Fell_Oxy; 1.
DR SMART; SM00702; P4HC; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
DR KW Hypothetical protein.
SQ SEQUENCE 736 AA; 83411 MW; EA1819938FAE685E CRC64;

Query Match 55.3%; Score 407; DB 4; Length 736;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 707; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAVRALKLLTLLAVAAASQAQVESEAGWMTPTDLLFAEGTAAYARGDWPVLSMER 60
DB 1 MAVRALKLLTLLAVAAASQAQVESEAGWMTPTDLLFAEGTAAYARGDWPVLSMER 60
QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120
DB 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120
QY 121 RRCGLGPPAAHSLSEEMLEFRKSPNYLOVAFKINKLEKAVAAAHFFVGNPEHMEMQ 180
DB 121 RRCGLGPPAAHSLSEEMLEFRKSPNYLOVAFKINKLEKAVAAAHFFVGNPEHMEMQ 180
QY 181 QNLDDYQTMGSKVEADFKDLETQPHMQEFLGLVRLYSEEQPQEAAPHLEAALQEFVAYE 240
DB 181 QNLDDYQTMGSKVEADFKDLETQPHMQEFLGLVRLYSEEQPQEAAPHLEAALQEFVAYE 240
QY 241 ECRALCEGPDYDGYNLYEYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPFEDFL 300
DB 241 ECRALCEGPDYDGYNLYEYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPFEDFL 300
QY 361 SAKERYQRSLLEKELLFFAYDFVFGIPFVDPDPSWTPEEVIKPLQKQKSERETAIRISQE 420
DB 361 SAKERYQRSLLEKELLFFAYDFVFGIPFVDPDPSWTPEEVIKPLQKQKSERETAIRISQE 420
QY 421 IGNLMKEIETLVBEKTKESLDYSRLTREGGPLLLEGISLTWNKLLNGVQRVMDGVISD 480
DB 421 IGNLMKEIETLVBEKTKESLDYSRLTREGGPLLLEGISLTWNKLLNGVQRVMDGVISD 480
QY 481 HECQBLRLTNVAATSGDGYRGQTSPHTPNEKFGYGVTVFKALKLQGEKQKVPQLSAHLYN 540
DB 481 HECQBLRLTNVAATSGDGYRGQTSPHTPNEKFGYGVTVFKALKLQGEKQKVPQLSAHLYN 540
QY 541 VTEKVRIMESYFRDLTPLYPSYSHLVCTAEEVQAEKDDSHPHVYDNCILNAETLVC 600
```

```
DB 541 VTEKVRIMESYFRDLTPLYPSYSHLVCTAEEVQAEKDDSHPHVYDNCILNAETLVC 600
QY 601 VKEPPAYTFRDYSAILYNGDFDGNFYFTELDATKTVAEQPQCGRVGFSSGTENPHG 660
DB 601 VKEPPAYTFRDYSAILYNGDFDGNFYFTELDATKTVAEQPQCGRVGFSSGTENPHG 660
QY 661 VKAVTRGRCALALWFTLDPHRSERDRVQADDLVKMLFSPREMDLSQSQP 710
DB 661 VKAVTRGRCALALWFTLDPHRSERDRVQADDLVKMLFSPREMDLSQSQP 710
```

## RESULT 4

```
Q9HC87
ID Q9HC87 PRELIMINARY; PRT; 363 AA.
AC Q9HC87;
DT 01-WAR-2001 (TrEMBLrel. 16, Created)
DT 01-WAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE GROS1-S protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kaul S.C., Sugihara T., Yoshida A., Nomura H., Wadhwa R.;
RT "Gros1, a potential growth suppressor on chromosome 1: its identity to
RT basement membrane-associated proteoglycan, leprecan."
RL Oncogene 19:3576-3583(2000).
DR EMBL; AF097431; AAG31018.1; -.
DR GO; GO:0008285; P:negative regulation of cell proliferation; NAS.
DR InterPro; IPR008941; TPR-like.
SQ SEQUENCE 363 AA; 41158 MW; 5B1B66B38679E76A CRC64;
```

```
Query Match 48.9%; Score 360; DB 4; Length 363;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MAVRALKLLTLLAVAAASQAQVESEAGWMTPTDLLFAEGTAAYARGDWPVLSMER 60
DB 1 MAVRALKLLTLLAVAAASQAQVESEAGWMTPTDLLFAEGTAAYARGDWPVLSMER 60
QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120
DB 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120
QY 121 RRCGLGPPAAHSLSEEMLEFRKSPNYLOVAFKINKLEKAVAAAHFFVGNPEHMEMQ 180
DB 121 RRCGLGPPAAHSLSEEMLEFRKSPNYLOVAFKINKLEKAVAAAHFFVGNPEHMEMQ 180
QY 181 QNLDDYQTMGSKVEADFKDLETQPHMQEFLGLVRLYSEEQPQEAAPHLEAALQEFVAYE 240
DB 181 QNLDDYQTMGSKVEADFKDLETQPHMQEFLGLVRLYSEEQPQEAAPHLEAALQEFVAYE 240
QY 241 ECRALCEGPDYDGYNLYEYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPFEDFL 300
DB 241 ECRALCEGPDYDGYNLYEYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPFEDFL 300
QY 301 PSHNYLQFAYNIGNYTOAGECAKTYLLFFPNDEVNQNLAAYAAMLGEEHTRSIGPRE 360
DB 301 PSHNYLQFAYNIGNYTOAGECAKTYLLFFPNDEVNQNLAAYAAMLGEEHTRSIGPRE 360
```

## RESULT 5

```
Q96BR8
ID Q96BR8 PRELIMINARY; PRT; 390 AA.
AC Q96BR8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
```





```
QY 661 VKAVTRGQRCALWFTLDRHSR 685
Db 661 VKAVTRGQRCALWFTLDRHSR 685

RESULT 7
Q9H6K3 PRELIMINARY; PRT; 390 AA.
AC Q9H6K3
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ22188.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK025841; BAB15256.1; -.
DR GO; GO:0016706; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0019538; P:protein metabolism; IEA.
DR InterPro; IPR005123; 2OG-Fell_Oxy.
DR InterPro; IPR00886; ER_target_S.
DR InterPro; IPR006620; Pro_4_hyd_alph.
DR Pfam; PF03171; 2OG-Fell_Oxy; 1.
DR SMART; SM00702; P4HC; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
DR KW Hypothetical protein_TARGT; 1.
SQ SEQUENCE 390 AA; 44093 MW; 61D29993108677FB CRC64;

Query Match 25.5%; Score 188; DB 4; Length 390;
Best Local Similarity 99.5%; Pred. No. 1.1e-182;
Matches 388; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 347 MLGSEHTRSGPRESKEYQSRILKELLFFAYDVGIPFVDPDSWTPPEVTPKRLQEK 406
Db 1 MLGSEHTRSGPRESKEYQSRILKELLFFAYDVGIPFVDPDSWTPPEVTPKRLQEK 60

QY 407 QKSERETAVRISOEIGNLMKEITLVVEKTKESLDVSRLTREGGPLYEGISLTWNSKLL 466
Db 61 QKSERETAVRISOEIGNLMKEITLVVEKTKESLDVSRLTREGGPLYEGISLTWNSKLL 120

QY 467 NGQVRVMDGVI SDHECQELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLQ 526
Db 121 NGQVRVMDGVI SDHECQELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLQ 180

QY 527 EGKVPQSAHLNVYTKVRIMESYRLDTPLFYSHLVCRTAEEVQAEKDDSHPV 586
Db 181 EGKVPQSAHLNVYTKVRIMESYRLDTPLFYSHLVCRTAEEVQAEKDDSHPV 240

QY 587 HVNDCILNAETLVCKEPPAYTRDYSAILYLNGDFDGGNFYFTELDKAKTVTAEVQPCG 646
Db 241 HVNDCILNAETLVCKEPPAYTRDYSAILYLNGDFDGGNFYFTELDKAKTVTAEVQPCG 300

QY 647 RAVGFSSGTENPHGVKAVTGQRCALWFTLDRHSRDRVQADLLVKMLFSPPEMDLS 706
Db 301 RAVGFSSGTENPHGVKAVTGQRCALWFTLDRHSRDRVQADLLVKMLFSPPEMDLS 360

QY 707 QEQLDAQQPPPEPAQESLSGSESKPKDEL 736
Db 361 QEQLDAQQPPPEPAQESLSGSESKPKDEL 390

RESULT 8
Q9RIJ8 PRELIMINARY; PRT; 728 AA.
QY 661 VKAVTRGQRCALWFTLDRHSR 685
Db 661 VKAVTRGQRCALWFTLDRHSR 685

Query Match 13.7%; Score 101; DB 11; Length 739;
Best Local Similarity 100.0%; Pred. No. 1.3e-93;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 602 KEPPAYTRDYSAILYLNGDFDGGNFYFTELDKAKTVTAEVQPCGRAVGFSSGTENPHGV 661
Db 594 KEPPAYTRDYSAILYLNGDFDGGNFYFTELDKAKTVTAEVQPCGRAVGFSSGTENPHGV 653

QY 662 KAVTRGQRCALWFTLDRHSRDRVQADLLVKMLFSPPE 702
Db 654 KAVTRGQRCALWFTLDRHSRDRVQADLLVKMLFSPPE 694

RESULT 9
Q8CFL3 PRELIMINARY; PRT; 739 AA.
ID Q8CFL3
AC Q8CFL3
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Similar to leprecan.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC024047; AAH24047.1; -.
DR PIR; P00551; PT0551.
DR GO; GO:0016706; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR InterPro; IPR005123; 2OG-Fell_Oxy.
DR InterPro; IPR00886; ER_target_S.
DR InterPro; IPR006620; Pro_4_hyd_alph.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; 2OG-Fell_Oxy; 1.
DR SMART; SM00702; P4HC; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
SQ SEQUENCE 739 AA; 83592 MW; 64B7AE568B0868EA CRC64;

Query Match 13.7%; Score 101; DB 11; Length 739;
Best Local Similarity 100.0%; Pred. No. 1.3e-93;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 602 KEPPAYTRDYSAILYLNGDFDGGNFYFTELDKAKTVTAEVQPCGRAVGFSSGTENPHGV 661
Db 594 KEPPAYTRDYSAILYLNGDFDGGNFYFTELDKAKTVTAEVQPCGRAVGFSSGTENPHGV 653

QY 662 KAVTRGQRCALWFTLDRHSRDRVQADLLVKMLFSPPE 702
Db 654 KAVTRGQRCALWFTLDRHSRDRVQADLLVKMLFSPPE 694

Query Match 13.7%; Score 101; DB 11; Length 728;
Best Local Similarity 100.0%; Pred. No. 1.3e-93;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 602 KEPPAYTRDYSAILYLNGDFDGGNFYFTELDKAKTVTAEVQPCGRAVGFSSGTENPHGV 661
Db 594 KEPPAYTRDYSAILYLNGDFDGGNFYFTELDKAKTVTAEVQPCGRAVGFSSGTENPHGV 653

QY 662 KAVTRGQRCALWFTLDRHSRDRVQADLLVKMLFSPPE 702
Db 654 KAVTRGQRCALWFTLDRHSRDRVQADLLVKMLFSPPE 694
```

```

Best Local Similarity 100.0%; Pred. No. 1.3e-93;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 602 KEPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAEVPQCGRAVFGSSGTENPHGV 661
Db 605 KEPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAEVPQCGRAVFGSSGTENPHGV 664

Qy 662 KAVTRGRCALWFTLDPHRSERDRVQADDLVKMLFSPEE 702
Db 665 KAVTRGRCALWFTLDPHRSERDRVQADDLVKMLFSPEE 705

RESULT 10
Q9CWX5 PRELIMINARY; PRT; 742 AA.
AC Q9CWX5;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE 2410024C15RIK protein.
GN LEPRE1 OR GROS1 OR 2410024C15RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryonic stem cells;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischnann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK010578; BAB27041.1; -.
DR MGD; MGI:1888921; Leprel.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0016049; P:cell growth; IDA.
DR InterPro; IPR005123; 2OG-FeII_Oxy.
DR InterPro; IPR005620; Pro 4 hvd_alph.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; 2OG-FeII_Oxy; 1.
DR SMART; SM00702; P4HC; 1.
SQ SEQUENCE 742 AA; 84086 MW; CFC44D34D96ACEA7 CRC64;

Query Match 13.7%; Score 101; DB 11; Length 742;
Best Local Similarity 100.0%; Pred. No. 1.3e-93;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 602 KEPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAEVPQCGRAVFGSSGTENPHGV 661
Db 605 KEPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAEVPQCGRAVFGSSGTENPHGV 664

Qy 662 KAVTRGRCALWFTLDPHRSERDRVQADDLVKMLFSPEE 702
Db 665 KAVTRGRCALWFTLDPHRSERDRVQADDLVKMLFSPEE 705

RESULT 11
Q9QZT7 PRELIMINARY; PRT; 747 AA.
AC Q9QZT7;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Growth suppressor 1L.
GN LEPRE1 OR GROS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD1 ICR; TISSUE=Fibroblast;
RX MEDLINE=20406537; PubMed=10951563;
RA Kaul S.C., Sugihara T., Yoshida A., Nomura H., Wadhwa R.;
RT "Gros1, a potential growth suppressor on chromosome 1: its identity to
RT basement membrane-associated proteoglycan, leprecan.";
RL Oncogene 19:3576-3583(2000).
DR EMBL; AF165163; AAF04806.1; -.
DR MGD; MGI:1888921; Leprel.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0016049; P:cell growth; IDA.
DR InterPro; IPR005123; 2OG-FeII_Oxy.
DR InterPro; IPR006620; Pro 4 hvd_alph.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; 2OG-FeII_Oxy; 1.
DR SMART; SM00702; P4HC; 1.
SQ SEQUENCE 747 AA; 84811 MW; EBEAB2A62E824123 CRC64;

Query Match 12.4%; Score 91; DB 11; Length 747;
Best Local Similarity 100.0%; Pred. No. 2.1e-83;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 612 YSAILYLNGDFDGNFYFTLDAKTVAEVPQCGRAVFGSSGTENPHGVKAVTRGRC 671
Db 614 YSAILYLNGDFDGNFYFTLDAKTVAEVPQCGRAVFGSSGTENPHGVKAVTRGRC 673

Qy 672 IALWFTLDPHRSERDRVQADDLVKMLFSPEE 702
Db 674 IALWFTLDPHRSERDRVQADDLVKMLFSPEE 704

RESULT 12
Q9QZT6 PRELIMINARY; PRT; 542 AA.
AC Q9QZT6;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Growth suppressor 1S.
GN LEPRE1 OR GROS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD1 ICR;
RX MEDLINE=20406537; PubMed=10951563;
RA Kaul S.C., Sugihara T., Yoshida A., Nomura H., Wadhwa R.;
RT "Gros1, a potential growth suppressor on chromosome 1: its identity to
RT basement membrane-associated proteoglycan, leprecan.";
RL Oncogene 19:3576-3583(2000).
DR EMBL; AF165164; AAF04807.1; -.
DR MGD; MGI:1888921; Leprel.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0016049; P:cell growth; IDA.

```

```
DR InterPro; IPR008941; TPR-like.
SQ SEQUENCE 542 AA; 61599 MW; 0A211411F71DF16B CRC64;

Query Match
Best Local Similarity 11.1%; Score 82; DB 11; Length 542;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 383 FGIPIFVDPDSWTPEEVIPIKLEQKQKSERETAVRISQEIQNLKMKETETIVESKTKESLDV 442
DB 385 FGIPIFVDPDSWTPEEVIPIKLEQKQKSERETAVRISQEIQNLKMKETETIVESKTKESLDV 444
QY 443 SRLTREGGPLYEGISLTMNSK 464
DB 445 SRLTREGGPLYEGISLTMNSK 466

RESULT 13
Q8BSV2 Q8BSV2 PRELIMINARY; PRT; 738 AA.
AC Q8BSV2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Leprecan 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=pituitary;
RA The FANTOM Consortium.
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK030436; BAC26962.1; -.
DR MGI; MGI:2146663; AW553532.
DR GO; GO:0016706; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0019538; P:protein metabolism; IEA.
DR InterPro; IPR005123; 2OG-Fell_Oxy.
DR InterPro; IPR00886; ER_target_S.
DR InterPro; IPR006620; Pro_4_hyd_alph.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; 2OG-Fell_Oxy; 1.
DR SMART; SM00702; P4HC; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 409 AA; 46325 MW; 49CB7CAA66F909FC CRC64;

Query Match
Best Local Similarity 2.2%; Score 16; DB 11; Length 409;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 656 ENPHGVKAVTRGQRC 671
DB 350 ENPHGVKAVTRGQRC 365

RESULT 15
Q8CG71 Q8CG71 PRELIMINARY; PRT; 703 AA.
AC Q8CG71;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Myxoid liposarcoma associated protein 4.
GN AW553532 OR MLAT4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL;
RA Thein-Jarnum S., Aman P.;
RT "The novel gene Myxoid Liposarcoma Associated Transcript 4 (MLAT4)
belongs to the Grosi/leprecan proteoglycan family."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ430350; CAD23038.1; -.
DR MGI; MGI:2146663; AW553532.
DR GO; GO:0016706; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0019538; P:protein metabolism; IEA.
DR InterPro; IPR005123; 2OG-Fell_Oxy.
DR InterPro; IPR00886; ER_target_S.
DR InterPro; IPR006620; Pro_4_hyd_alph.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; 2OG-Fell_Oxy; 1.
DR SMART; SM00702; P4HC; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
SQ SEQUENCE 703 AA; 80154 MW; 1F42F9B9938573E4 CRC64;

Query Match
Best Local Similarity 2.2%; Score 16; DB 11; Length 703;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
DR InterPro; IPR008941; TPR-like.
SQ SEQUENCE 542 AA; 61599 MW; 0A211411F71DF16B CRC64;

Query Match
Best Local Similarity 11.1%; Score 82; DB 11; Length 542;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 383 FGIPIFVDPDSWTPEEVIPIKLEQKQKSERETAVRISQEIQNLKMKETETIVESKTKESLDV 442
DB 385 FGIPIFVDPDSWTPEEVIPIKLEQKQKSERETAVRISQEIQNLKMKETETIVESKTKESLDV 444
QY 443 SRLTREGGPLYEGISLTMNSK 464
DB 445 SRLTREGGPLYEGISLTMNSK 466

RESULT 13
Q8BSV2 Q8BSV2 PRELIMINARY; PRT; 738 AA.
AC Q8BSV2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Leprecan 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=pituitary;
RA The FANTOM Consortium.
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK030436; BAC26962.1; -.
DR MGI; MGI:2146663; AW553532.
DR GO; GO:0016706; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0019538; P:protein metabolism; IEA.
DR InterPro; IPR005123; 2OG-Fell_Oxy.
DR InterPro; IPR00886; ER_target_S.
DR InterPro; IPR006620; Pro_4_hyd_alph.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; 2OG-Fell_Oxy; 1.
DR SMART; SM00702; P4HC; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
SQ SEQUENCE 738 AA; 83528 MW; 75967DE318F55F4A CRC64;

Query Match
Best Local Similarity 10.9%; Score 80; DB 11; Length 738;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 602 KEPPAYTRDYSAIYLINGDFDGGNFYETLDAKTVTAEVQPCGGRVGFSSGTENPHGV 661
DB 604 KEPPAYTRDYSAIYLINGDFDGGNFYETLDAKTVTAEVQPCGGRVGFSSGTENPHGV 663
QY 662 KAVTRGQRCATLWFTLDP 681
DB 664 KAVTRGQRCATLWFTLDP 683

RESULT 14
Q8C673 Q8C673 PRELIMINARY; PRT; 409 AA.
AC Q8C673;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
GN AW553532.
OS Mus musculus (Mouse).
```

Oy 656 ENPHGVKAVTRGQCA 571  
| | | | | | | | | | | | | | | |  
Db 644 ENPHGVKAVTRGQCA 659

Search completed: July 18, 2004, 09:59:46  
Job time : 54 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 18:12:26 ; Search time 148 Seconds  
(without alignments)  
2759.756 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 736  
Sequence: 1 MAVRALKLLTLLAVVAAS.....PPEPAQESUSGESKPKDEL 736

Scoring table:  
OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Pgapop 6.0 , Pgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1360453

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool\_P/US10045815/runat\_14072004\_123119\_17267/app\_query.fasta\_1.903  
-DB=Issued Patents NA -SUFFIX=oli.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USR=US10045815@cgn\_1\_128@runat\_14072004\_123119\_17267 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTCUS COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	9	1.2	13856	4	US-08-956-171E-31
C 2	9	1.2	37769	4	US-08-311-731A-23
C 3	9	1.2	4403765	3	US-09-103-840A-2
C 4	9	1.2	441529	3	US-09-103-840A-1
C 5	8	1.1	251	4	US-08-956-171E-4554
C 6	8	1.1	301	4	US-09-439-313-271
C 7	8	1.1	301	4	US-09-352-616A-271
C 8	8	1.1	301	4	US-09-232-149A-271
C 9	8	1.1	301	4	US-09-159-812-271
C 10	8	1.1	301	4	US-09-636-215-271
C 11	8	1.1	301	4	US-09-685-166A-271
C 12	8	1.1	301	4	US-09-688-489-271

C 13	8	1.1	363	4	US-08-407-620A-12	Sequence 12, Appl
C 14	8	1.1	375	4	US-09-252-991A-4191	Sequence 4191, Ap
C 15	8	1.1	387	4	US-09-582-337-21	Sequence 21, Appl
C 16	8	1.1	408	4	US-09-075-460-14	Sequence 14, Appl
C 17	8	1.1	480	3	US-09-039-555B-11	Sequence 11, Appl
C 18	8	1.1	646	4	US-09-614-912-135	Sequence 135, App
C 19	8	1.1	657	4	US-09-252-991A-4339	Sequence 4339, Ap
C 20	8	1.1	663	4	US-09-252-991A-3252	Sequence 3252, Ap
C 21	8	1.1	789	4	US-09-252-991A-4083	Sequence 4083, Ap
C 22	8	1.1	855	4	US-09-252-991A-4107	Sequence 4107, Ap
C 23	8	1.1	894	4	US-09-328-352-1260	Sequence 1260, Ap
C 24	8	1.1	987	4	US-09-221-017B-1115	Sequence 1115, Ap
C 25	8	1.1	1105	4	US-09-221-017B-108	Sequence 108, App
C 26	8	1.1	1155	4	US-09-252-991A-4515	Sequence 4515, Ap
C 27	8	1.1	1281	4	US-09-252-991A-3691	Sequence 3691, Ap
C 28	8	1.1	1515	4	US-09-252-991A-4123	Sequence 4123, Ap
C 29	8	1.1	1602	4	US-09-252-991A-3776	Sequence 3776, Ap
C 30	8	1.1	1608	2	US-08-211-718-8	Sequence 8, Appli
C 31	8	1.1	1755	4	US-09-252-991A-3588	Sequence 3588, Ap
C 32	8	1.1	1842	4	US-09-489-039A-2888	Sequence 2888, Ap
C 33	8	1.1	2004	1	US-08-471-033-18	Sequence 18, Appl
C 34	8	1.1	2004	2	US-08-471-044-18	Sequence 18, Appl
C 35	8	1.1	2004	2	US-08-463-483A-18	Sequence 18, Appl
C 36	8	1.1	2004	2	US-08-471-046A-18	Sequence 18, Appl
C 37	8	1.1	2004	2	US-08-470-566B-18	Sequence 18, Appl
C 38	8	1.1	2004	2	US-08-469-334-18	Sequence 18, Appl
C 39	8	1.1	2004	3	US-09-300-529-18	Sequence 18, Appl
C 40	8	1.1	2576	1	US-08-471-033-35	Sequence 35, Appl
C 41	8	1.1	2576	2	US-08-471-044-35	Sequence 35, Appl
C 42	8	1.1	2576	2	US-08-463-483A-35	Sequence 35, Appl
C 43	8	1.1	2576	2	US-08-471-046A-35	Sequence 35, Appl
C 44	8	1.1	2576	2	US-08-470-566B-35	Sequence 35, Appl
C 45	8	1.1	2576	2	US-08-469-334-35	Sequence 35, Appl

ALIGNMENTS

RESULT 1

US-08-956-171E-31/c  
; Sequence 31, Application US/08956171E  
; Patent No. 6593114

GENERAL INFORMATION:  
APPLICANT: Charles Kunsch

Gil H. Choi  
Patrick S. Dillon  
Craig A. Rosen  
Steven C. Barash  
Michael R. Fannon

TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 5256  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/956,171E  
FILING DATE: 20-Oct-1997

CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/009,861  
FILING DATE: January 5, 1996

APPLICATION NUMBER: 08/781,986  
FILING DATE: January 3, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mark J. Hyman



```
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.: 4,696-04 Length: 4411529
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.22% Indels: 0
DB: 3 Gaps: 0

US-10-045-815-4 (1-736) x US-09-103-840A-1 (1-4411529)
QY 14 AlavalavalalalalaserGlnAla 22
Db 3093432 GCCGTGTCGCCGCCGACCCAGCGC 3093458

RESULT 5
US-08-956-171E-4554
; Sequence 4554, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 4554:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 251 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 4554:
US-08-956-171E-4554

Alignment Scores:
Pred. No.: 39.6 Length: 251
Score: 8.00 Matches: 8

; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-10-045-815-4 (1-736) x US-08-956-171E-4554 (1-251)
QY 723 GluSerLeuSerGlySerGluSer 730
Db 152 GAATCGCTATCTGGATCTGAGTCG 175

RESULT 6
US-09-439-313-271/c
; Sequence 271, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqui
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 271
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(301)
; OTHER INFORMATION: n = A,T,C or G
US-09-439-313-271

Alignment Scores:
Pred. No.: 47.2 Length: 301
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-09-439-313-271 (1-301)
QY 673 AlaLeuTrpPheThrLeuAspPro 680
Db 193 GCTCTGTGGTTCACTTGACCCA 170

RESULT 7
US-09-352-616A-271/c
; Sequence 271, Application US/09352616A
; Patent No. 6395278
; GENERAL INFORMATION:
; APPLICANT: Dillon, Davin C.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang, Yuqui
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.427C8
; CURRENT APPLICATION NUMBER: US/09/352,616A
; CURRENT FILING DATE: 1999-07-13
```

```

; NUMBER OF SEQ ID NOS: 472
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 271
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(301)
; OTHER INFORMATION: n = A,T,C or G
US-09-352-616A-271

Alignment Scores:
Pred. No.: 47.2 Length: 301
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-09-352-616A-271 (1-301)

Qy 673 AlaleuTrpPheThrLeuAspPro 680
Db 193 GCTCTGGTTCACCTGGACCCA 170

RESULT 8
US-09-232-149A-271/c
; Sequence 271, Application US/09232149A
; Patent No. 6465611
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE
; FILE REFERENCE: 210121.427C6
; CURRENT APPLICATION NUMBER: US/09/232.149A
; CURRENT FILING DATE: 1999-01-15
; NUMBER OF SEQ ID NOS: 338
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 271
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapien
; NAME/KEY: misc_feature
; LOCATION: (1)...(301)
; OTHER INFORMATION: n = A,T,C or G
US-09-232-149A-271

Alignment Scores:
Pred. No.: 47.2 Length: 301
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-09-232-149A-271 (1-301)

Qy 673 AlaleuTrpPheThrLeuAspPro 680
Db 193 GCTCTGGTTCACCTGGACCCA 170

RESULT 9
US-09-159-812-271/c
; Sequence 271, Application US/09159812A
; Patent No. 6613872
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF
; NUMBER OF SEQ ID NOS: 472
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 271
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(301)
; OTHER INFORMATION: n = A,T,C or G
US-09-636-215-271

Alignment Scores:
Pred. No.: 47.2 Length: 301
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-09-636-215-271 (1-301)

Qy 673 AlaleuTrpPheThrLeuAspPro 680
Db 193 GCTCTGGTTCACCTGGACCCA 170

RESULT 10
US-09-636-215-271/c
; Sequence 271, Application US/09636215
; Patent No. 6620922
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.42717C17
; CURRENT APPLICATION NUMBER: US/09/636.215
; CURRENT FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 852
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 271
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(301)
; OTHER INFORMATION: n = A,T,C or G
US-09-636-215-271

Alignment Scores:
Pred. No.: 47.2 Length: 301
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0

```

; TITLE OF INVENTION: PROSTATE CANCER AND METHODS FOR THEIR USE

; FILE REFERENCE: 210121.428C5

; CURRENT APPLICATION NUMBER: US/09/159,812A

; CURRENT FILING DATE: 1998-09-23

; NUMBER OF SEQ ID NOS: 306

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 271

; LENGTH: 301

; TYPE: DNA

; ORGANISM: Homo sapien

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(301)

; OTHER INFORMATION: n = A,T,C or G

US-09-159-812-271

Alignment Scores:

Pred. No.: 47.2 Length: 301

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 1.09% Indels: 0

DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-09-159-812-271 (1-301)

Qy 673 AlaleuTrpPheThrLeuAspPro 680

Db 193 GCTCTGGTTCACCTGGACCCA 170

RESULT 10

US-09-636-215-271/c

; Sequence 271, Application US/09636215

; Patent No. 6620922

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Dillon, Davin C.

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Harlocker, Susan L.

; APPLICANT: Jiang, Yuqui

; APPLICANT: Henderson, Robert A.

; APPLICANT: Kalos, Michael D.

; APPLICANT: Fanger, Gary R.

; APPLICANT: Retter, Marc W.

; APPLICANT: Stolk, John A.

; APPLICANT: Day, Craig H.

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Carter, Darrick

; APPLICANT: Li, Samuel

; APPLICANT: Wang, Aijun

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Hepler, William

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE REFERENCE: 210121.42717C17

; CURRENT APPLICATION NUMBER: US/09/636.215

; CURRENT FILING DATE: 2000-08-10

; NUMBER OF SEQ ID NOS: 852

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 271

; LENGTH: 301

; TYPE: DNA

; ORGANISM: Homo sapien

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(301)

; OTHER INFORMATION: n = A,T,C or G

US-09-636-215-271

Alignment Scores:

Pred. No.: 47.2 Length: 301

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0



Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.09% Indels: 0  
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-09-636-215-271 (1-301)

Qy 673 AlaLeuTrpPheThrLeuAspPro 680  
Db 193 GCTCTGTGGTTCACCTTGGACCCA 170

RESULT 11

US-09-685-166A-271/c  
; Sequence 271, Application US/09685166A  
; Patent No. 6630305  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Harlocker, Susan L.  
; APPLICANT: Jiang, Yuqi  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Stolk, John A.  
; APPLICANT: Day, Craig H.  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Carter, Darrick  
; APPLICANT: Li, Samuel  
; APPLICANT: Wang, Aijun  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Hepler, William  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
; FILE REFERENCE: 210121.427C21  
; CURRENT APPLICATION NUMBER: US/09/685,166A  
; CURRENT FILING DATE: 2000-10-10  
; NUMBER OF SEQ ID NOS: 898  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 271  
; LENGTH: 301  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(301)  
; OTHER INFORMATION: n = A,T,C or G

US-09-685-166A-271

Alignment Scores:  
Pred. No.: 47.2 Length: 301  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.09% Indels: 0  
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-09-685-166A-271 (1-301)

Qy 673 AlaLeuTrpPheThrLeuAspPro 680  
Db 193 GCTCTGTGGTTCACCTTGGACCCA 170

RESULT 12

US-09-688-489-271/c  
; Sequence 271, Application US/09688489  
; Patent No. 6664377  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer Lynn  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE  
; TITLE OF INVENTION: CANCER AND METHODS FOR THEIR USE

; FILE REFERENCE: 210121.427D2  
; CURRENT APPLICATION NUMBER: US/09/688,489  
; CURRENT FILING DATE: 2000-10-13  
; NUMBER OF SEQ ID NOS: 338  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 271  
; LENGTH: 301  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(301)  
; OTHER INFORMATION: n = A,T,C or G

US-09-688-489-271

Alignment Scores:  
Pred. No.: 47.2 Length: 301  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.09% Indels: 0  
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-09-688-489-271 (1-301)

Qy 673 AlaLeuTrpPheThrLeuAspPro 680  
Db 193 GCTCTGTGGTTCACCTTGGACCCA 170

RESULT 13

US-08-407-620A-12/c  
; Sequence 12, Application US/08407620A  
; Patent No. 6569430  
; GENERAL INFORMATION:  
; APPLICANT: WALDMANN, HERMAN  
; APPLICANT: CLARK, MICHAEL R.  
; APPLICANT: WINTER, GREGORY P.  
; APPLICANT: RIECHMANN, LUTZ  
; TITLE OF INVENTION: ANTIBODIES  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHUYE P.C.  
; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: U.S.A.  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/407,620A  
; FILING DATE: 21-MAR-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/235,705  
; FILING DATE: 29-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/099,480  
; FILING DATE: 30-JUL-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/921,601  
; FILING DATE: 03-AUG-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/424,233  
; FILING DATE: 12-OCT-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 88036228  
; FILING DATE: 12-FEB-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 8804464

```
/ FILING DATE: 25-FEB-1988
/ ATTORNEY/AGENT INFORMATION:
/ NAME: MITCHARD, LEONARD C.
/ REGISTRATION NUMBER: 29,009
/ REFERENCE/DOCKET NUMBER: 604-325
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (703) 816-4000
/ TELEFAX: (703) 816-4100
/ TELEX: 200797 NIXN UR
/ INFORMATION FOR SEQ ID NO: 12:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 363 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
US-08-407-620A-12

Alignment Scores:
Pred. No.: 56.5 Length: 363
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-08-407-620A-12 (1-363)
QY 718 ProGluProAlaGlnGluSerLeu 725
Db 77 CCAGAACCTGCACAGAGAGCTC 54

RESULT 14
US-09-252-991A-4191
/ Sequence 4191, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ PRIOR FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 4191
/ LENGTH: 375
/ TYPE: DNA
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4191

Alignment Scores:
Pred. No.: 58.3 Length: 375
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-09-252-991A-4191 (1-375)
QY 60 ArgAlaLeuArgSerArgAlaAla 67
Db 212 CGGCGCTTCGATCAGCGCGCG 235

RESULT 15
US-09-582-337-21/C
/ Sequence 21, Application US/09582337
/ Patent No. 6562618
/ GENERAL INFORMATION:
/ APPLICANT: Japan Tobacco, Inc.
/ TITLE OF INVENTION: Monoclonal Antibody Against Connective Tissue Growth Factor
/ FILE REFERENCE: J1-009PCT
/ CURRENT APPLICATION NUMBER: US/09/582,337
/ CURRENT FILING DATE: 2000-06-23
/ PRIOR APPLICATION NUMBER: JP P1997-367699
/ PRIOR FILING DATE: 1997-12-25
/ PRIOR APPLICATION NUMBER: JP P1998-356183
/ PRIOR FILING DATE: 1998-12-15
/ NUMBER OF SEQ ID NOS: 27
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 21
/ LENGTH: 387
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(387)
/ NAME/KEY: sig_peptide
/ LOCATION: (1)..(47)
/ OTHER INFORMATION: Initiation codon and a portion of a signal
/ OTHER INFORMATION: sequence are lacked.
/ NAME/KEY: V region
/ LOCATION: (48)..(335)
US-09-582-337-21

Alignment Scores:
Pred. No.: 60.1 Length: 387
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x US-09-582-337-21 (1-387)
QY 651 PheSerSerGlyThrGluAsnPro 658
Db 50 TTTCATCCGCGACCGAAACCCA 27

Search completed: July 19, 2004, 00:22:40
Job time : 3383 secs
```

Result No.	Score	Query			DB	ID	Description
		Match	Length				
1	9	1.2	75	1	TRB3_AGRTU	P54912 agrobacteri	
2	9	1.2	364	1	DRG2_HUMAN	P55039 homo sapien	
3	9	1.2	364	1	DRG2_MOUSE	Q9qxb9 mus musculu	
4	8	1.1	162	1	CAV2_FUGRU	Q9ygm9 fugu rubrip	
5	8	1.1	401	1	CASP_HUMAN	Q75718 homo sapien	
6	8	1.1	431	1	SC65_RAT	Q64375 rattus norv	
7	8	1.1	437	1	N055_HUMAN	Q92791 homo sapien	
8	8	1.1	493	1	UPEI_NPVCD	Q90158 choristoneu	
9	8	1.1	508	1	OTOL_NCCKE	P83371 oncorhynch	
10	7	1.0	84	1	IML3_ORYSA	Q9XGYS oryza sativ	
11	7	1.0	103	1	YKN7_YEAST	P36064 saccharomyc	
12	7	1.0	121	1	TGFA_MACMU	P55244 macaca mula	
13	7	1.0	126	1	RL7_CHLTE	Q8kg16 chlorobium	
14	7	1.0	133	1	TGPA_SHEEP	P98135 ovis aries	
15	7	1.0	159	1	TGFA_MOUSE	P48030 mus musculu	
16	7	1.0	159	1	TGFA_RAT	P01134 rattus norv	
17	7	1.0	160	1	TGFA_HUMAN	P01135 homo sapien	
18	7	1.0	160	1	TGFA_PIG	Q06922 sus scrofa	
19	7	1.0	183	1	HSIV_XANCP	Q8p551 xanthomonas	
20	7	1.0	196	1	RR4_ALOPR	P36445 alopecurus	
21	7	1.0	196	1	R84_ANDIS	P36446 andropogon	
22	7	1.0	196	1	R84_ANTOD	P36447 anthoxanthu	
23	7	1.0	196	1	R84_ARUJA	P36448 arundinaria	
24	7	1.0	196	1	R84_BRAPI	P36449 brachypodi	
25	7	1.0	196	1	R84_BROER	P36451 bromus erect	
26	7	1.0	196	1	R84_CALEP	P36452 calamagrost	
27	7	1.0	196	1	R84_DENGI	P36454 dendrocalam	
28	7	1.0	196	1	R84_ELEIN	P36456 eleusine in	
29	7	1.0	196	1	R84_ELYCA	P36457 elymus cana	
30	7	1.0	196	1	R84_FESGI	P36458 festuca gra	
31	7	1.0	196	1	R84_FESPR	P36459 festuca pra	
32	7	1.0	196	1	R84_LYGPS	P36464 lygeum spar	
33	7	1.0	196	1	R84_MELAL	P36465 melica mali	

100

Db 155 QRSLEKEL 163

RESULT 4

CAV2\_FUGRU STANDARD; PRT; 162 AA.

AC Q9YGM9;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Caveolin-2.

GN CAV2 OR CAV-2.

OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Notoleosteii;

OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;

OC Tetraodontidae; Tetraodontidae; Takifugu.

OX NCBI\_TaxID=31033;

[1]

RN SEQUENCE FROM N.A.

RP Cottage A.J.;

RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.

[2]

RN SEQUENCE FROM N.A.

RX MEDLINE=22799194; PubMed=12917688;

RA Thomas J.W., Touchman J.W., Blakesley R.W., Bouffard G.G., Beckstrom-Sternberg S.M., Margulies E.H., Blanchette M., Siepel A.C., Thomas P.J., McDowell J.C., Maskeri B., Hansen N.F., Schwartz M.S., Weber R.J., Kent W.J., Karolchik D., Bruen T.C., Bevan R., Cutler D.J., Schwartz S., Elaitaki L., Idol J.R., Prasad A.B., Lee-Lin S.-Q., Maduro V.V., Summers T.J., Portnoy M.E., Dietrich N.L., Akhter N., Ayele K., Benjamin B., Cariaga K., Brinkley C.P., Brooks S.Y., Granits E., Guan X., Gupta J., Haghighi P., Ho S.-L., Huang M.C., Karlins E., Laric P.L., Legaspi R., Lim M.J., Maduro Q.L., Masinello C.A., Mastrian S.D., McCloskey J.C., Pearson R., Stanitrop S., Tingson E.E., Tran J.T., Tsugeon C., Vogt J.L., Walker M.A., Wetherby K.D., Wiggins L.S., Young A.C., Zhang L.-H., Ooegawa K., Zhu B., Zhao B., Shu C.L., De Jong P.J., Lawrence C.E., Smit A.F., Chakravarti A., Haussler D., Green P., Miller W., Green E.D.;

RT "Comparative analyses of multi-species sequences from targeted genomic regions.";

RL Nature 424:788-793(2003).

CC -!- FUNCTION: May act as a scaffolding protein within caveolar membranes. Interacts directly with G-protein alpha subunits and can functionally regulate their activity (By similarity).

CC -!- SUBUNIT: Homooligomer (By similarity).

CC -!- SUBCELLULAR LOCATION: Membrane protein of caveolae. Potential hairpin-like structure in the membrane (By similarity).

CC -!- SIMILARITY: Belongs to the caveolin family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AJ010316; CAA09081.1; -

DR EMBL; AC090119; AAL40363.1; -

DR InterPro; IPR001612; Caveolin.

DR Pfam; PF01146; Caveolin; 1.

DR PROSITE; PS01210; CAVROLIN; 1.

KW Transmembrane; Lipoprotein.

FT DOMAIN 1 86 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 87 107 POTENTIAL.

FT DOMAIN 108 162 CYTOPLASMIC (POTENTIAL).

SQ SEQUENCE 162 AA; 18236 MW; 1D7CF4907D491253 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 162;

Best Local Similarity 100.0%; Pred. No. 4.1;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LITLLAV 15

Db 87 LITLLAV 94

RESULT 5

CASP\_HUMAN STANDARD; PRT; 401 AA.

ID CASP\_HUMAN

AC O757I8;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Cartilage-associated protein precursor.

GN CRTAP OR CASP

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

[1]

RN SEQUENCE FROM N.A.

RP TISSUE=Petal brain;

RC MEDLINE=20169181; PubMed=10702664;

RX Tonachini L., Morello R., Monticone M., Skaug J., Scherer S.W., Cancedda R., Castagnola P.;

RA "cDNA cloning, characterization and chromosome mapping of the gene encoding human cartilage associated protein (CRTAP).";

RT Cytogenet. Cell Genet. 87:191-194(1999).

RL [2]

RN SEQUENCE FROM N.A.

RP TISSUE=Kidney;

RC MEDLINE=22388257; PubMed=12477932;

RX Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.W., Hong L., Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullihy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fates J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -!- SUBCELLULAR LOCATION: Secreted. Extracellular matrix (By similarity).

CC -!- TISSUE SPECIFICITY: Found in articular chondrocytes. Expressed in a variety of tissues

CC -!- SIMILARITY: BELONGS TO THE CRTAP / NO55 FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AJ006470; CAA07054.1; -

DR EMBL; BC008745; AAO8745.1; -

DR Genew; HGNC:2379; CRTAP.

DR MIM; 605497; -

DR InterPro; IPR008940; Prenyl trans.

KW Extracellular matrix; Signal.

FT SIGNAL 1 POTENTIAL.

```

FT CHAIN 27 401 CARTILAGE-ASSOCIATED PROTEIN.
FT CARBOHYD 87 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 363 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CONFLICT 137 E -> D (IN REF. 2).
SQ SEQUENCE 401 AA; 46561 MW; 4BEEED4089195456F CRC64;

Query Match 1.1%; Score 8; DB 1; Length 401;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 8; Conservative 0; Mismatches 0; Gaps 0;

Qy 110 FGGLLRRA 117
| | | | |
Db 109 FGGLLRRA 116

RESULT 6
SC65 RAT STANDARD; PRT; 431 AA.
AC Q64375;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Synaptonemal complex protein SC65.
GN SC65.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
TISSUE=Testis;
RX MEDLINE=932113429; PubMed=1363622;
RA Chen Q., Pearlman R.E., Moens P.B.;
RT "Isolation and characterization of a cDNA encoding a synaptonemal
complex protein.";
RL Biochem. Cell Biol. 70:1030-1038(1992).
CC -!- SUBCELLULAR LOCATION: Nuclear. Located in the pairing zone of the
synaptonemal complex.
CC -!- TISSUE SPECIFICITY: Found in testis, brain, heart and at a much
lower level in liver.
CC -!- SIMILARITY: BELONGS TO THE CRTAP / NO55 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; X65454; CAA46449.1; -.
CC F01; A56822; A56822.
CC InterPro; IPR008940; Prenyl_trans.
KW Nuclear protein.
FT DOMAIN 319 328 GLU-RICH (ACIDIC).
FT DOMAIN 353 384 GLU-RICH (ACIDIC).
SQ SEQUENCE 431 AA; 49995 MW; 4A34F3029407B2E7 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 431;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 116 RAACLRRC 123
| | | | |
Db 85 RAACLRRC 92

RESULT 7
NO55 HUMAN STANDARD; PRT; 437 AA.
AC Q92791; Q9H4F6;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)

```

```

DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Nuclear autoantigen No55.
GN SC65 OR NO55.
OC Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
TISSUE=Bladder epithelium;
RX MEDLINE=97015880; PubMed=8862517;
RA Ochs R.L., Stein T.W. Jr., Chan E.K.I., Ruutu M., Tan E.M.;
RT "cDNA cloning and characterization of a novel nucleolar protein.";
RL Mol. Biol. Cell 7:1015-1024(1996).
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE=20411121; PubMed=10952778;
RA Fossa A., Siebert R., Aasheim H.C., Maelandsmo G.M., Berner A.,
RA Fossa S.D., Paus E., Smeland E.B., Gaudernack G.;
RT "Identification of nucleolar protein No55 as a tumour-associated
autoantigen in patients with prostate cancer.";
RL Br. J. Cancer 83:743-749(2000).
RN [3]
SEQUENCE FROM N.A.
TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar. Localized uniformly
throughout the granular component of the nucleolus and on the
surface of chromosomes during mitosis.
CC -!- SIMILARITY: BELONGS TO THE CRTAP / NO55 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; U47621; AAC51792.1; -.
CC EMBL; AJ250583; CAC16786.1; -.
CC EMBL; BC001047; AAH01047.1; -.
CC EMBL; BC007942; AAH07942.1; -.
CC EMBL; BC011701; AAH11701.1; -.
CC GK; Q92791.
CC GO; GO:0000230; C:nuclear mitotic chromosome; TAS.
CC GO; GO:0005730; C:nucleolus; TAS.
CC GO; GO:0005716; C:synaptonemal complex; TAS.
CC GO; GO:0007130; P:synaptonemal complex formation; TAS.
CC InterPro; IPR008940; Prenyl_trans.
CC Nuclear protein; Antigen.
CC DOMAIN 346 437 ASP/GLU-RICH (ACIDIC).
CC CONFLICT 186 186 Q -> R (IN REF. 2).

```

```

GN OTOL.
OS Oncorhynchus keta (Chum salmon).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8018;
RN [1]
SEQUENCE FROM N.A., SEQUENCE OF 26-40 AND 452-482, FUNCTION, AND
TISSUE SPECIFICITY.
RP MEDLINE=21845944; PubMed=11856329;
RX Murayama E., Takagi Y., Ohita T., Davis J.G., Greene M.I.,
RA Nagasawa H.;
RA "Fish otolith contains a unique structural protein, otolin-1.";
ET Eur. J. Biochem. 269:688-696(2002).
RL -1- FUNCTION: May be part of the internal framework of the otolith
CC where it may provide nucleation sites to facilitate
CC calcification.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Selectively expressed in the sacculus where it
CC is localised to the otolith, the gelatinous layer of the otolithic
CC membrane, and part of the transitional epithelium.
CC -1- PTM: N-glycosylated.
CC -1- SIMILARITY: Contains 1 collagenous domain.
CC -1- SIMILARITY: Contains 1 Clq domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; AB067770; BAB84561.1; -.
DR GO; GO:0005576; C:extracellular; IDA.
DR GO; GO:0045299; P:otolith mineralization; NAS.
DR InterPro; IPR001073; Clq.
DR InterPro; IPR008161; Clq_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008983; TNF_like.
DR Pfam; PF00386; Clq; 1.
DR Pfam; PF01391; Collagen; 4.
DR PRINTS; PR00007; COMPLEMENTC1Q.
DR PRODOM; PD000007; Clq_helix; 1.
DR PROSITE; PS01113; Clq; 1.
DR Collagen; Glycoprotein; Signal.
FT SIGNAL 1 25
FT CHAIN 26 508 OTOLIN-1.
FT DOMAIN 144 367 COLLAGEN-LIKE.
FT FT DOMAIN 369 505 Clq.
FT FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT CARBOHYD 416 416 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 508 AA; 52138 MW; 90438799ACE3E9BE CRC64;

Query Match 1.1%; Score 8; DB 1; Length 508;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 LITLAVV 16
DB 10 LITLAVV 17
|||||||

RESULT 10
IM13 ORYSA STANDARD; PRT; 84 AA.
AC Q9XG5;
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 10-OCT-2003 (rel. 42, Last annotation update)
DE Mitochondrial import inner membrane translocase subunit Tim13.
GN TIM13.
OS Oryza sativa (Rice).

```





FT PROPEP 2 16 REMOVED IN MATURE FORM (BY SIMILARITY).  
FT CHAIN 17 66 TRANSFORMING GROWTH FACTOR ALPHA.  
FT PROPEP 67 121 REMOVED IN MATURE FORM (BY SIMILARITY).  
FT DOMAIN 2 75 EXTRACELLULAR (BY SIMILARITY).  
FT TRANSMEM 76 101 BY SIMILARITY.  
FT DOMAIN 20 60 EGF-LIKE (BY SIMILARITY).  
FT DISULFID 24 37 BY SIMILARITY.  
FT DISULFID 32 48 BY SIMILARITY.  
FT DISULFID 50 59 BY SIMILARITY.  
FT CARBOHYD 3 3 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT NON TER 121 121  
SQ SEQUENCE 121 AA; 13182 MW; 463031DFBEC14816 CRC64;  
  
Query Match 1.0%; Score 7; DB 1; Length 121;  
Best Local Similarity 100.0%; Pred. No. 34; Mismatches 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0;  
  
QY 12 LLA VAA 18  
| | | | |  
DB 64 LLA VAA 70  
  
RESULT 13  
RL7 CHLITE STANDARD; PRT; 126 AA.  
AC O8KG16;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE S05 ribosomal protein L7/L12.  
GN RPL7 OR CTO154.  
OS Chlorobium tepidum.  
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;  
OC Chlorobium.  
OX NCBI\_TaxID=1097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TLS / ATCC 49652 / DSM 12025;  
RX MEDLINE=22103685; PubMed=12093901;  
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,  
RA Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Kolonay J.L., Yang F.,  
RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Parksey D.,  
RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.F., Parksey D.,  
RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,  
RA Vamathevan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,  
RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;  
RT "The complete genome sequence of Chlorobium tepidum TLS, a  
photosynthetic anaerobic, green-sulfur bacterium";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).  
CC -!- FUNCTION: Seems to be the binding site for several of the factors  
involved in protein synthesis and appears to be essential for  
accurate translation (By similarity).  
CC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE012794; AAM71402.1; -  
DR TIGR; CT0154; -  
DR HAMAP; MF\_00368; -; 1.  
DR InterPro; IPR008932; Ribos L12/7 olig.  
DR InterPro; IPR002026; Ribosomal L12.  
DR Pfam; PF00542; Ribosomal L12; 1.  
DR ProDom; PD001326; Ribosomal L12; 1.  
DR Trifams; TIGR00855; L12; 1.  
DR Ribosomal protein; Complete proteome.  
KW Ribosomal protein; Complete proteome.  
SQ SEQUENCE 126 AA; 12996 MW; 31CFB56AC94FE054 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 126;  
Best Local Similarity 100.0%; Pred. No. 35; Mismatches 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0;  
  
QY 428 IETLVEE 434  
| | | | |  
DB 4 IETLVEE 10  
  
RESULT 14  
TGFA SHEEP STANDARD; PRT; 133 AA.  
ID TGFA SHEEP  
AC P98135;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Transforming growth factor alpha precursor (TGF-alpha) (EGF-like TGF)  
DE (ETGF) (TGF type 1) (Fragment).  
GN TGFA OR TGF-A.  
OS Ovis aries (Sheep).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Caprinae; Ovis.  
OX NCBI\_TaxID=9940;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Merino; TISSUE=Skin dorsal;  
RX MEDLINE=95268698; PubMed=7749621;  
RA Sutton R., Ward W.G., Raphael K.A., Cam G.R.;  
RT "Growth factor expression in skin during wool follicle development";  
RL Comp. Biochem. Physiol. 110B:697-705(1995).  
CC -!- FUNCTION: TGF alpha is a mitogenic polypeptide that is able to  
bind to the EGF receptor and to act synergistically with TGF beta  
to promote anchorage-independent cell proliferation in soft agar.  
CC -!- SUBUNIT: Interacts with the PDZ domains of SDCBP and SNTA1. The  
interaction with SDCBP, is required for the targeting to the cell  
surface (By similarity).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (precursor form);  
extracellular (mature form).  
CC -!- TISSUE SPECIFICITY: Skin.  
CC -!- DEVELOPMENTAL STAGE: Wool follicle development.  
CC -!- SIMILARITY: Contains 1 EGF-like domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; L36232; AAA53113.1; -  
DR HSSP; P01135; 3TGF.  
DR InterPro; IPR001336; EGF 1.  
DR InterPro; IPR006209; EGF-like.  
DR InterPro; IPR006210; IEGF.  
DR PRINTS; PRO0009; EGFTGF.  
DR SMART; SM00181; EGF; 1.  
DR PROSITE; PS00022; EGF 1; 1.  
DR PROSITE; PS01186; EGF 2; 1.  
DR PROSITE; PS00026; EGF 3; 1.  
KW Growth factor; Mitogen; Glycoprotein; EGF-like domain; Transmembrane;  
KW Signal.  
FT SIGNAL 1 23 POTENTIAL.  
FT PROPEP 24 38 REMOVED IN MATURE FORM.  
FT CHAIN 39 88 TRANSFORMING GROWTH FACTOR ALPHA.  
FT PROPEP 89 >133 REMOVED IN MATURE FORM.  
FT DOMAIN 24 97 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 98 120 POTENTIAL.  
FT DOMAIN 121 >133 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 42 82 EGF-LIKE.  
FT DISULFID 46 59 BY SIMILARITY.  
FT DISULFID 54 70 BY SIMILARITY.

FT DISULFID 72 81 BY SIMILARITY.  
FT CARBOHYD 25 25 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT NON TER 133 133  
SQ SEQUENCE 133 AA; 14026 MW; F9F8E03BAA28AFB1 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 133;  
Best Local Similarity 100.0%; Pred.No. 37;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LLAVVAA 18  
DB 86 LLAVVAA 92

## RESULT 15

TGFA\_MOUSE  
ID TGFA\_MOUSE STANDARD; PRT; 159 AA.  
AC P48030;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Transforming growth factor alpha precursor (TGF-alpha) (EGF-like TGF)  
DE (ETGF) (TGF type 1).  
GN TGFA.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. PubMed=1420315;  
RA MEDLINE=93041937; PubMed=8871107;  
RA Vaughan T.J., Pascall J.C., Brown K.D.;  
RT "Nucleotide sequence and tissue distribution of mouse transforming growth factor-alpha";  
RL Biochim. Biophys. Acta 1132:322-324(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SV; TISSUE=Brain;  
RX MEDLINE=97031181; PubMed=8871107;  
RA Berkowitz E.A., Serogy K.B., Schroeder J.A., Russell W.E.,  
RA Evans E.P., Riedel R.F., Phillips H.K., Harrison C.A., Lee D.C.,  
RA Luetke N.C.;  
RT "Characterization of the mouse transforming growth factor alpha gene: its expression during eyelid development and in waved 1 tissues.";  
RL Cell Growth Differ. 7:1271-1282(1996).  
CC -!- FUNCTION: TGF alpha is a mitogenic polypeptide that is able to bind to the EGF receptor and to act synergistically with TGF beta to promote anchorage-independent cell proliferation in soft agar.  
CC -!- SUBUNIT: Interacts with the PDZ domains of SDCBP and SNTAL. The interaction with SDCBP is required for the targeting to the cell surface (By similarity).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (precursor form); extracellular (mature form).  
CC -!- SIMILARITY: Contains 1 EGF-like domain.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; M92420; AAA040481.1; -  
CC EMBL; U65016; AAB50554.1; -  
CC EMBL; U64873; AAB50553.1; -  
CC PIR; S27195; S27195.  
CC PDB; 1GK5; 08-AUG-02.  
CC MGD; MGI:98724; Tgfa.  
CC InterPro; IPR001336; EGF 1.  
CC InterPro; IPR006209; EGF-like.  
CC InterPro; IPR006210; IEGF.  
CC Pfam; PF00008; EGF; 1.

DR PRINTS: PR00009; EGTFGF.  
DR SMART: SM00181; EGF; 1.  
DR PROSITE; PS00022; EGF\_1; 1.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS00026; EGF\_3; 1.  
KW Growth factor; Mitogen; Glycoprotein; EGF-like domain; Transmembrane; Signal; Lipoprotein; Palmitate; 3D-structure.  
FT SIGNAL 1 23  
FT PROPEP 24 38  
FT CHAIN 39 88  
FT PROPEP 89 159  
FT DOMAIN 24 97  
FT TRANSMEM 98 123  
FT DOMAIN 124 159  
FT DOMAIN 44 83  
FT DISULFID 46 59  
FT DISULFID 54 70  
FT DISULFID 72 81  
FT LIPID 152 152  
FT LIPID 153 153  
FT CARBOHYD 25 25  
SQ SEQUENCE 159 AA; 17018 MW; BE0AE8D9CE7DDFD CRC64;

Query Match 1.0%; Score 7; DB 1; Length 159;  
Best Local Similarity 100.0%; Pred.No. 43;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LLAVVAA 18  
DB 86 LLAVVAA 92

Search completed: July 18, 2004, 09:58:40  
Job time : 18 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 09:57:04 ; Search time 23 Seconds  
(without alignments)  
1652.031 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 736  
Sequence: 1 MAVRALKLTLLVVAAS.....PPEPAQESUGSEKPKDEL 736

Scoring table:  
OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/prodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/prodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/prodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/prodata/2/iaa/6C COMB.pep.\*  
6: /cgn2\_6/prodata/2/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	1.1	220	4	US-09-252-991A-19823
2	8	1.1	426	4	US-09-252-991A-20262
3	8	1.1	493	3	US-08-942-012B-28
4	8	1.1	613	4	US-09-489-039A-10059
5	7	1.0	50	1	US-08-208-008C-7
6	7	1.0	75	4	US-09-621-976-5127
7	7	1.0	82	1	US-07-847-743B-16
8	7	1.0	82	1	US-08-456-201-16
9	7	1.0	82	2	US-08-330-161-14
10	7	1.0	82	2	US-08-456-241-16
11	7	1.0	82	2	US-08-440-401-14
12	7	1.0	82	2	US-08-419-878B-14
13	7	1.0	82	4	US-09-173-480-14
14	7	1.0	82	5	PCT-US92-04295A-16
15	7	1.0	87	2	US-07-885-089B-12
16	7	1.0	115	3	US-08-545-809A-122
17	7	1.0	124	4	US-09-489-039A-11363
18	7	1.0	125	4	US-09-489-039A-10706
19	7	1.0	159	4	US-09-227-853A-14
20	7	1.0	171	3	US-08-611-587-5
21	7	1.0	177	4	US-09-252-991A-29834
22	7	1.0	184	4	US-09-252-991A-32941
23	7	1.0	185	4	US-09-540-236-3570
24	7	1.0	189	4	US-09-252-991A-18839
25	7	1.0	202	4	US-09-543-681A-7996
26	7	1.0	210	3	US-08-611-587-4
27	7	1.0	211	4	US-09-252-991A-26393

28 7 1.0 219 4 US-08-311-731A-305 Sequence 305, App  
29 7 1.0 243 4 US-09-252-991A-19011 Sequence 19011, A  
30 7 1.0 274 4 US-09-489-039A-13000 Sequence 13000, A  
31 7 1.0 290 4 US-09-134-001C-3401 Sequence 3401, Ap  
32 7 1.0 297 4 US-09-584-568C-7 Sequence 7, Appli  
33 7 1.0 302 4 US-09-252-991A-30761 Sequence 30761, A  
34 7 1.0 316 4 US-09-711-184-378 Sequence 378, App  
35 7 1.0 325 4 US-09-252-991A-18010 Sequence 18010, A  
36 7 1.0 345 1 US-08-843-993-1 Sequence 1, Appli  
37 7 1.0 345 3 US-09-059-520A-1 Sequence 1, Appli  
38 7 1.0 345 3 US-09-334-275-1 Sequence 1, Appli  
39 7 1.0 351 4 US-09-540-236-3819 Sequence 3819, Ap  
40 7 1.0 352 4 US-09-252-991A-17906 Sequence 17906, A  
41 7 1.0 352 4 US-09-328-352-6547 Sequence 6547, Ap  
42 7 1.0 357 4 US-09-252-991A-28480 Sequence 28480, A  
43 7 1.0 364 4 US-09-242-859A-12 Sequence 12, Appl  
44 7 1.0 387 3 US-08-968-563-37 Sequence 37, Appl  
45 7 1.0 387 3 US-08-969-683A-37 Sequence 37, Appl

ALIGNMENTS

RESULT 1  
US-09-252-991A-19823  
; Sequence 19823, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 19823  
; LENGTH: 220  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-19823

Query Match 1.1%; Score 8; DB 4; Length 220;  
Best Local Similarity 100.0%; Pred. No. 17; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0;  
QY 96 AQASGAGA 103  
Db 7 AQASGAGA 14

RESULT 2  
US-09-252-991A-20262  
; Sequence 20262, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 20262  
; LENGTH: 426  
; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-20262

Query Match 1.1%; Score 8; DB 4; Length 426;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 RRAACLR 122  
| | | | |  
DB 206 RRAACLR 213

RESULT 3  
US-08-942-012B-28

; Sequence 28, Application US/08942012B  
; Patent No. 6235278  
; GENERAL INFORMATION:  
; APPLICANT: Miller, Lois K.  
; APPLICANT: Lu, Albert  
; APPLICANT: Dierks, Peter  
; APPLICANT: Black, Bruce  
; TITLE OF INVENTION: Biological Insect Control Agents Expressing  
; TITLE OF INVENTION: Insect-Specific Toxin Genes, Methods and Compositions  
; FILE REFERENCE: 28-96a  
; CURRENT APPLICATION NUMBER: US/08/942,012B  
; CURRENT FILING DATE: 1997-10-01  
; PRIOR APPLICATION NUMBER: 08/729,606  
; PRIOR FILING DATE: 2000-10-01  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 28  
; LENGTH: 493  
; TYPE: PRT  
; ORGANISM: Cf defective nucleopolyhedrosis virus  
US-08-942-012B-28

Query Match 1.1%; Score 8; DB 3; Length 493;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LLTLLAV 15  
| | | | |  
DB 5 LLTLLAV 12

RESULT 4  
US-09-489-039A-10059

; Sequence 10059, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 10059  
; LENGTH: 613  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-10059

Query Match 1.1%; Score 8; DB 4; Length 613;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 SQAEVESE 27  
| | | | |  
DB 407 SQAEVESE 414

RESULT 5  
US-08-208-008C-7

; Sequence 7, Application US/08208008C  
; Patent No. 5633147  
; GENERAL INFORMATION:  
; APPLICANT: MEISSNER, ET AL.  
; TITLE OF INVENTION: Transforming Growth Factor  
; TITLE OF INVENTION: Alpha - H1  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/208,008C  
; FILING DATE: March 8, 1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MULLINS, J.G.  
; REGISTRATION NUMBER: 33,073  
; REFERENCE/DOCKET NUMBER: 325800-98  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 50 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-08-208-008C-7

Query Match 1.0%; Score 7; DB 1; Length 50;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LLAVVAA 18  
| | | | |  
DB 43 LLAVVAA 49

RESULT 6

US-09-621-976-5127  
; Sequence 5127, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 5127  
; LENGTH: 75  
; TYPE: PRT  
; ORGANISM: Homo sapiens

FEATURE:  
NAME/KEY: UNSURE  
LOCATION: 62  
OTHER INFORMATION: Xaa = Gly,Ser  
NAME/KEY: UNSURE  
LOCATION: 69  
OTHER INFORMATION: Xaa = Ile,Thr  
US-09-621-976-5127

Query Match 1.0%; Score 7; DB 4; Length 75;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 176 HMEQQN 182  
Db 54 HMEQQN 60

RESULT 7  
US-07-847-743B-16  
Sequence 16, Application US/07847743B  
Patent No. 5367060  
GENERAL INFORMATION:  
APPLICANT: Genentech, Inc.  
TITLE OF INVENTION: Structure, Production and Use of  
TITLE OF INVENTION: Heregulin  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/847,743B  
FILING DATE: 19920306  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/705256  
FILING DATE: 24-May-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/765212  
FILING DATE: 25-Sep-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/790801  
FILING DATE: 08-No. 5367060-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Hensley, Max D.  
REGISTRATION NUMBER: 27,043  
REFERENCE/DOCKET NUMBER: 712P3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/266-1489  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 82 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear

Query Match 1.0%; Score 7; DB 1; Length 82;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 LLAVAA 18  
Db 12 LLAVAA 18

Db 43 LLAVAA 49.

RESULT 8  
US-08-456-201-16  
Sequence 16, Application US/08456201  
Patent No. 5641869  
GENERAL INFORMATION:  
APPLICANT: Vandlen, Richard L.,  
TITLE OF INVENTION: Structure, Production and Use of  
TITLE OF INVENTION: Heregulin 2 Ligands  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/456,201  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/126,145  
FILING DATE:  
APPLICATION NUMBER: 07/880,917  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/705256  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/765212  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/790801  
ATTORNEY/AGENT INFORMATION:  
NAME: Hensley, Max D.  
REGISTRATION NUMBER: 27,034  
REFERENCE/DOCKET NUMBER: 712P4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/266-1489  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 82 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-456-201-16

Query Match 1.0%; Score 7; DB 1; Length 82;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 LLAVAA 18  
Db 43 LLAVAA 49

RESULT 9  
US-08-330-161-14  
Sequence 14, Application US/08330161  
Patent No. 5834229  
GENERAL INFORMATION:  
APPLICANT: Vandlen, Richard  
APPLICANT: Holmes, William  
TITLE OF INVENTION: Structure, Production and Use of Heregulin 2 Ligands  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Genentech, Inc.  
 STREET: 460 Point San Bruno Blvd  
 CITY: South San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94080  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: patin (Genentech)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/330,161  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/035430  
 FILING DATE: 22-MAR-1993  
 APPLICATION NUMBER: 07/705256  
 FILING DATE: 24-MAY-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Adler, Carolyn R.  
 REGISTRATION NUMBER: 32,324  
 REFERENCE/DOCKET NUMBER: 712C1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415/225-2614  
 TELEFAX: 415/952-9881  
 TELEX: 910/371-7168  
 INFORMATION FOR SEQ ID NO: 14:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 82 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 US-08-330-161-14

Query Match 1.0%; Score 7; DB 2; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 65;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LLAVVAA 18  
 |||||  
 Db 43 LLAVVAA 49

RESULT 10  
 US-08-456-241-16  
 ; Sequence 16, Application US/08456241  
 ; Patent No. 5840525  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Vandlen, Richard L.  
 ; APPLICANT: Holmes, William E.  
 ; TITLE OF INVENTION: STRUCTURE, PRODUCTION AND USE OF  
 ; TITLE OF INVENTION: HREGULIN  
 ; NUMBER OF SEQUENCES: 30  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genentech, Inc.  
 ; STREET: 460 Point San Bruno Blvd  
 ; CITY: South San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94080  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: patin (Genentech)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/456,241  
 ; FILING DATE: 31-MAY-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/126145  
 ; FILING DATE: 23-SEP-1993

;; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/880917  
 ; FILING DATE: 11-MAY-1992  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/847743  
 ; FILING DATE: 06-MAR-1992  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/790801  
 ; FILING DATE: 08-NOV-1991  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/765212  
 ; FILING DATE: 25-SEP-1991  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/705256  
 ; FILING DATE: 24-MAY-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Lee, Wendy M.  
 ; REGISTRATION NUMBER: 00,000  
 ; REFERENCE/DOCKET NUMBER: 712P4C1D1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 415/225-1994  
 ; TELEFAX: 415/952-9881  
 ; TELEX: 910/371-7168  
 ; INFORMATION FOR SEQ ID NO: 16:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 82 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; US-08-456-241-16

Query Match 1.0%; Score 7; DB 2; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 65;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LLAVVAA 18  
 |||||  
 Db 43 LLAVVAA 49

RESULT 11  
 US-08-440-401-14  
 ; Sequence 14, Application US/08440401  
 ; Patent No. 5856110  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Vandlen, Richard L.  
 ; APPLICANT: Holmes, William E.  
 ; TITLE OF INVENTION: Structure, Production and Use of  
 ; TITLE OF INVENTION: Heregulin 2 Ligands  
 ; NUMBER OF SEQUENCES: 17  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genentech, Inc.  
 ; STREET: 460 Point San Bruno Blvd  
 ; CITY: South San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94080  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: patin (Genentech)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/440,401  
 ; FILING DATE: 12-MAY-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/330161  
 ; FILING DATE: 25-OCT-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/035430  
 ; FILING DATE: 22-MAR-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/705256

FILING DATE: 24-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER: 00,000  
REFERENCE/DOCKET NUMBER: 712C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1994  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 82 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-440-401-14

Query Match 1.0%; Score 7; DB 2; Length 82;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LLAVVAA 18  
Db 43 LLAVVAA 49

RESULT 13  
US-09-173-480-14  
Sequence 14, Application US/09173480  
Patent No. 6399746  
GENERAL INFORMATION:  
APPLICANT: Vandlen, Richard L.  
ADDRESSEE: Holmes, William E.  
TITLE OF INVENTION: Structure, Production and Use of Heregulin 2 Ligands  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/173,480  
FILING DATE: 14-Oct-1998  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/440,401  
FILING DATE: <Unknown>  
APPLICATION NUMBER: 08/035430  
FILING DATE: 22-MAR-1993  
APPLICATION NUMBER: 07/705256  
FILING DATE: 24-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER: 00,000  
REFERENCE/DOCKET NUMBER: 712C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1994  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 82 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
US-09-173-480-14

Query Match 1.0%; Score 7; DB 4; Length 82;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LLAVVAA 18  
Db 43 LLAVVAA 49

RESULT 14  
PCT-US92-04295A-16  
Sequence 16, Application PC/TUS9204295A  
GENERAL INFORMATION:  
APPLICANT: Genentech, Inc.  
TITLE OF INVENTION: Structure, Production and Use of

FILING DATE: 24-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER: 00,000  
REFERENCE/DOCKET NUMBER: 712C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1994  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 82 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-440-401-14

Query Match 1.0%; Score 7; DB 2; Length 82;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LLAVVAA 18  
Db 43 LLAVVAA 49

RESULT 12  
US-08-419-878B-14  
Sequence 14, Application US/08419878B  
Patent No. 5859206  
GENERAL INFORMATION:  
APPLICANT: Vandlen, Richard L.  
ADDRESSEE: Holmes, William E.  
TITLE OF INVENTION: Antibodies Specific For Heregulin 2-alpha  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/419,878B  
FILING DATE: 11-Apr-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/330161  
FILING DATE: 25-OCT-1994  
APPLICATION NUMBER: 08/035430  
FILING DATE: 22-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER: 40,378  
REFERENCE/DOCKET NUMBER: P0712C2D1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-1994  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 82 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
US-08-419-878B-14

;/ TITLE OF INVENTION: Heregulin  
;/ NUMBER OF SEQUENCES: 30  
;/ CORRESPONDENCE ADDRESS:  
;/ ADDRESSEE: Genentech, Inc.  
;/ STREET: 460 Point San Bruno Blvd  
;/ CITY: South San Francisco  
;/ STATE: California  
;/ COUNTRY: USA  
;/ ZIP: 94080  
;/ COMPUTER READABLE FORM:  
;/ MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
;/ COMPUTER: IBM PC compatible  
;/ OPERATING SYSTEM: PC-DOS/MS-DOS  
;/ SOFTWARE: patin (Genentech)  
;/ CURRENT APPLICATION DATA:  
;/ APPLICATION NUMBER: PCT/US92/04295A  
;/ FILING DATE: 19920521  
;/ CLASSIFICATION:  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER: 07/705256  
;/ FILING DATE: 24-MAY-1991  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER: 07/765212  
;/ FILING DATE: 25-SEP-1991  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER: 07/790801  
;/ FILING DATE: 08-NOV-1991  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER: 07/847743  
;/ FILING DATE: 06-MAR-1992  
;/ ATTORNEY/AGENT INFORMATION:  
;/ NAME: Hensley, Max D.  
;/ REGISTRATION NUMBER: 27,043  
;/ REFERENCE/DOCKET NUMBER: 712P4  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: 415/225-1994  
;/ TELEFAX: 415/352-9881  
;/ TELEX: 910/371-7168  
;/ INFORMATION FOR SEQ ID NO: 16:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 82 amino acids  
;/ TYPE: AMINO ACID  
;/ TOPOLOGY: linear  
;/ PCT-US92-04295A-16

Query Match 1.0%; Score 7; DB 5; Length 82;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LLAVVAA 18  
Db 43 LLAVVAA 49

RESULT 15  
US-07-885-089B-12  
; Sequence 12, Application US/07885089B  
; Patent No. 5830595  
; GENERAL INFORMATION:  
; APPLICANT: Shoyab, Mohammed  
; APPLICANT: McDonald, Vicki L.  
; APPLICANT: Bradley, James G.  
; APPLICANT: Plowman, Gregory D.  
; TITLE OF INVENTION: AMPHIREGULINS: A FAMILY OF  
; TITLE OF INVENTION: HEPARIN-BINDING EPITHELIAL CELL GROWTH FACTORS  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036

;/ COMPUTER READABLE FORM:  
;/ MEDIUM TYPE: Floppy disk  
;/ COMPUTER: IBM PC compatible  
;/ OPERATING SYSTEM: PC-DOS/MS-DOS  
;/ SOFTWARE: PatentIn Release #1.0, Version #1.30  
;/ CURRENT APPLICATION DATA:  
;/ APPLICATION NUMBER: US/07/885,089B  
;/ FILING DATE: 18-MAY-1992  
;/ CLASSIFICATION: 530  
;/ ATTORNEY/AGENT INFORMATION:  
;/ NAME: Coruzzi, Laura A.  
;/ REGISTRATION NUMBER: 30,742  
;/ REFERENCE/DOCKET NUMBER: 5624-174  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: 212-790-9090  
;/ TELEFAX: 212-869-9741  
;/ TELEX: 66141 PENNIE  
;/ INFORMATION FOR SEQ ID NO: 12:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 87 amino acids  
;/ TYPE: amino acid  
;/ STRANDEDNESS: single  
;/ TOPOLOGY: linear  
;/ MOLECULE TYPE: peptide  
;/ US-07-885-089B-12

Query Match 1.0%; Score 7; DB 2; Length 87;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LLAVVAA 18  
Db 41 LLAVVAA 47

Search completed: July 18, 2004, 10:00:58  
Job time : 24 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 09:58:14 ; Search time 53 Seconds  
(without alignments)  
4340.465 Million cell updates/sec

Title: US-10-045-815-4

Perfect score: 736

Sequence: 1 MAVRALKLLTLLAVVAAS.....PPPAQESLSGSBSKPKDEL 736

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1285345 seqs, 312560633 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1285345

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	736	100.0	736	13	US-10-045-815-4
2	433	58.8	736	12	US-10-257-174-26
3	433	58.8	736	12	US-10-312-352-33
4	360	48.9	363	13	US-10-045-815-2
5	275	37.4	708	12	US-10-257-174-25
6	219	29.8	359	11	US-09-833-245-1251
7	101	13.7	173	11	US-09-833-245-1250
8	91	12.4	747	13	US-10-045-815-6
9	82	11.1	542	13	US-10-045-815-8
10	13	1.8	219	11	US-09-764-875-815
11	13	1.8	267	9	US-09-764-868-736
12	13	1.8	287	11	US-09-764-875-1106
13	13	1.8	282	11	US-10-012-697-1493
14	13	1.8	708	15	US-10-094-749-2560
15	8	1.1	10	10	US-09-572-404B-1740

#### ALIGNMENTS

##### RESULT 1

US-10-045-815-4  
; Sequence 4, Application US/10045815  
; Publication No. US20020160498A1  
; GENERAL INFORMATION:  
; APPLICANT: Wadhwa, Renu  
; APPLICANT: Sugihara, Takashi  
; APPLICANT: Ohide, Akiko  
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
; FILE REFERENCE: 06501-091001  
; CURRENT APPLICATION NUMBER: US/10/045,815  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: PCT/JP00/02731  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: JP 11/118806  
; PRIOR FILING DATE: 1999-04-26  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 736  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-045-815-4

Query Match	100.0%	Score 736:	DB 13;	Length 736;
Best Local Similarity	100.0%;	Pred.No. 0;		
Matches 736;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	NAVRALKLLTLLAVVAASQAQVESEAGWGMVTPDLLFAEGTAAYARGDWPGVVLMSER	60	
Db	1	NAVRALKLLTLLAVVAASQAQVESEAGWGMVTPDLLFAEGTAAYARGDWPGVVLMSER	60	
QY	61	ALRSPAALRALRCRTCAADFFWELDPDWPSPQAASGAGALRDLSPFGGLLRAACL	120	
Db	61	ALRSPAALRALRCRTCAADFFWELDPDWPSPQAASGAGALRDLSPFGGLLRAACL	120	
QY	121	RRCLGPPAAHSLSBEMELEFRKSPYNYLVAYFKINKLEKAVAAATFFVGNPEHMEMQ	180	

Sequence 2821, Ap  
Sequence 51334, A  
Sequence 64452, A  
Sequence 54435, A  
Sequence 79, Appl  
Sequence 6269, Ap  
Sequence 4501, Ap  
Sequence 143391,  
Sequence 2, Appli  
Sequence 64808, A  
Sequence 137691,  
Sequence 3528, Ap  
Sequence 8528, Ap  
Sequence 2863, Ap  
Sequence 9979, Ap  
Sequence 24, Appl  
Sequence 185927,  
Sequence 51318, A  
Sequence 544, App  
Sequence 32716, A  
Sequence 163956,  
Sequence 179700,  
Sequence 8355, Ap  
Sequence 11430, A  
Sequence 223454,  
Sequence 4597, Ap  
Sequence 133363,  
Sequence 8464, Ap  
Sequence 142593,  
Sequence 198322,

Db 121 RRLGPPAAHSLSEMELEFRKSPYNYLQVAFKINKLEKAVAAAHTFPVGNPEHMEMQ 180  
 Qy 181 QNLDYYQTMGSKVKEADFKDLETPHMQEFLGVRLVLYSEEOQOAVPHLEAALQOEFVAYE 240  
 Db 181 QNLDYYQTMGSKVKEADFKDLETPHMQEFLGVRLVLYSEEOQOAVPHLEAALQOEFVAYE 240  
 Qy 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTELASHSRKPEDFL 300  
 Db 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTELASHSRKPEDFL 300  
 Qy 301 PSHYNYLQFAYNIGNYTQAGECAKTYLLFFPNDEVNQNLAYAAMLGEEHTRSIGPRE 360  
 Db 301 PSHYNYLQFAYNIGNYTQAGECAKTYLLFFPNDEVNQNLAYAAMLGEEHTRSIGPRE 360  
 Qy 361 SAKYRQRSLLLEKELLFFAYDVFGIPFVDPDSWTPPEEVI PKRLOEKOKSERETAVRISQ 420  
 Db 361 SAKYRQRSLLLEKELLFFAYDVFGIPFVDPDSWTPPEEVI PKRLOEKOKSERETAVRISQ 420  
 Qy 421 IGNLMKEIETLVEEKTESLDVSLTRREGGPLYEGISLTWN SKLNGYQVVMGVID 480  
 Db 421 IGNLMKEIETLVEEKTESLDVSLTRREGGPLYEGISLTWN SKLNGYQVVMGVID 480  
 Qy 481 HECQELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLQBGKVPLOSALHYN 540  
 Db 481 HECQELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLQBGKVPLOSALHYN 540  
 Qy 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPHVDNCILNAETLVC 600  
 Db 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPHVDNCILNAETLVC 600  
 Qy 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAEYVQPCGKRAVFGSSGTENPHG 660  
 Db 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAEYVQPCGKRAVFGSSGTENPHG 660  
 Qy 661 VKAVTRGORCAIALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQOPLDAQGPPPEP 720  
 Db 661 VKAVTRGORCAIALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQOPLDAQGPPPEP 720  
 Qy 721 AQESLSGSESKEPKDEL 736  
 Db 721 AQESLSGSESKEPKDEL 736

RESULT 2

US-10-257-174-26  
 ; Sequence 26, Application US/10257174  
 ; Publication No. US20040034194A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Agarwal, Pankaj  
 ; APPLICANT: Murdoch, Paul R.  
 ; APPLICANT: Rizvi, Safia K.  
 ; APPLICANT: Smith, Randall F.  
 ; APPLICANT: Xiang, Zhaoying  
 ; TITLE OF INVENTION: NOVEL COMPOUNDS  
 ; FILE REFERENCE: GP50022  
 ; CURRENT APPLICATION NUMBER: US/10/257,174  
 ; CURRENT FILING DATE: 2002-10-10  
 ; PRIOR APPLICATION NUMBER: PCT/US01/11797  
 ; PRIOR FILING DATE: 2001-04-11  
 ; PRIOR APPLICATION NUMBER: 60/196,603  
 ; PRIOR FILING DATE: 2000-04-13  
 ; PRIOR APPLICATION NUMBER: 60/199,417  
 ; PRIOR FILING DATE: 2000-04-24  
 ; NUMBER OF SEQ ID NOS: 48  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 26  
 ; LENGTH: 736  
 ; TYPE: PRT  
 ; ORGANISM: Hmo sapiens  
 US-10-257-174-26

Query Match 58.8%; Score 433; DB 12; Length 736;  
 Best Local Similarity 99.6%; Pred. No. 0;

Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 MAVRALKLLTLLAVAAASQAEVSEAGMGWTPDLLFAEGTAAAYARGDWPGVVLWNER 60  
 Db 1 MAVRALKLLTLLAVAAASQAEVSEAGMGWTPDLLFAEGTAAAYARGDWPGVVLWNER 60  
 Qy 61 ALRSPAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRRAACL 120  
 Db 61 ALRSPAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRRAACL 120  
 Qy 121 RRLCGPPAAHSLSEMELEFRKSPYNYLQVAFKINKLEKAVAAAHTFPVGNPEHMEMQ 180  
 Db 121 RRLCGPPAAHSLSEMELEFRKSPYNYLQVAFKINKLEKAVAAAHTFPVGNPEHMEMQ 180  
 Qy 181 QNLDYYQTMGSKVKEADFKDLETPHMQEFLGVRLVLYSEEOQOAVPHLEAALQOEFVAYE 240  
 Db 181 QNLDYYQTMGSKVKEADFKDLETPHMQEFLGVRLVLYSEEOQOAVPHLEAALQOEFVAYE 240  
 Qy 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTELASHSRKPEDFL 300  
 Db 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTELASHSRKPEDFL 300  
 Qy 301 PSHYNYLQFAYNIGNYTQAGECAKTYLLFFPNDEVNQNLAYAAMLGEEHTRSIGPRE 360  
 Db 301 PSHYNYLQFAYNIGNYTQAGECAKTYLLFFPNDEVNQNLAYAAMLGEEHTRSIGPRE 360  
 Qy 361 SAKYRQRSLLLEKELLFFAYDVFGIPFVDPDSWTPPEEVI PKRLOEKOKSERETAVRISQ 420  
 Db 361 SAKYRQRSLLLEKELLFFAYDVFGIPFVDPDSWTPPEEVI PKRLOEKOKSERETAVRISQ 420  
 Qy 421 IGNLMKEIETLVEEKTESLDVSLTRREGGPLYEGISLTWN SKLNGYQVVMGVID 480  
 Db 421 IGNLMKEIETLVEEKTESLDVSLTRREGGPLYEGISLTWN SKLNGYQVVMGVID 480  
 Qy 481 HECQELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLQBGKVPLOSALHYN 540  
 Db 481 HECQELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLQBGKVPLOSALHYN 540  
 Qy 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPHVDNCILNAETLVC 600  
 Db 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPHVDNCILNAETLVC 600  
 Qy 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAEYVQPCGKRAVFGSSGTENPHG 660  
 Db 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAEYVQPCGKRAVFGSSGTENPHG 660  
 Qy 661 VKAVTRGORCAIALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQOPLDAQGPPPEP 720  
 Db 661 VKAVTRGORCAIALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQOPLDAQGPPPEP 720  
 Qy 721 AQESLSGSESKEPKDEL 736  
 Db 721 AQESLSGSESKEPKDEL 736

RESULT 3

US-10-312-352-33  
 ; Sequence 33, Application US/10312352  
 ; Publication No. US20040053824A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: INCYTE GENOMICS, INC.; TANG, Y. Tom  
 ; APPLICANT: YUE, Henry; AZIMZAI, Yalda  
 ; APPLICANT: HE, Ann; BATRA, Sajeev  
 ; APPLICANT: LO, Terence P.; NGUYEN, Dannie B.  
 ; APPLICANT: BURELL, John D.; MARCUS, Gregory A.  
 ; APPLICANT: ZINGLER, Kurt A.; GANDHI, Ameen R.  
 ; APPLICANT: LAL, Preeti G.; KEARNEY, Liam  
 ; APPLICANT: BURFORD, Neil; YAO, Monique G.  
 ; APPLICANT: CHAWLA, Narinder K.; ELLIOT, Vicki S.  
 ; APPLICANT: FARVIZU, Chandra S.; KHAN, Farrah A.  
 ; APPLICANT: BAUGHN, Mariah R.; HAFALIA, April, J.A.  
 ; APPLICANT: POLICKY, Jennifer L.; AU-YOUNG, Janice K.  
 ; APPLICANT: LU, Yan; BOROWSKY, Mark L.

APPLICANT: LU, Dyung Aina M.; RAMKUMAR, Javalaxmi  
APPLICANT: YANG, Juming; GURURAJAN, Rajagopal  
APPLICANT: WARREN, Bridget A.; GIETZEN, Kimberly J.  
APPLICANT: XU, Yuming; KALBUCK, Deborah A.  
APPLICANT: LEE, Ernestine A.; THANGAVELU, Kavitha  
APPLICANT: DELEGANE, Angelo M.; LEE, Sally  
TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES  
FILE REFERENCE: PF-0794 USN  
CURRENT APPLICATION NUMBER: US/10/312,352  
CURRENT FILING DATE: 2002-12-18  
PRIOR APPLICATION NUMBER: PCT/US01/21067  
PRIOR FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: US 60/215,454  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 60/219,462  
PRIOR FILING DATE: 2000-07-18  
PRIOR APPLICATION NUMBER: US 60/240,111  
PRIOR FILING DATE: 2000-10-12  
PRIOR APPLICATION NUMBER: US 60/240,106  
PRIOR FILING DATE: 2000-10-12  
PRIOR APPLICATION NUMBER: US 60/244,021  
PRIOR FILING DATE: 2000-10-27  
PRIOR APPLICATION NUMBER: US 60/248,887  
PRIOR FILING DATE: 2000-11-14  
PRIOR APPLICATION NUMBER: US 60/249,570  
PRIOR FILING DATE: 2000-11-16  
NUMBER OF SEQ ID NOS: 72  
SOFTWARE: PERL Program  
SEQ ID NO 33  
LENGTH: 736  
TYPE: PRT  
ORGANISM: Homo sapiens  
NAME/KEY: misc feature  
FEATURE:  
OTHER INFORMATION: Incyte ID No. US20040053824A1 6780147CD1  
US-10-312-352-33

Query Match 58.8%; Score 433; DB 12; Length 736;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MAVRALKLLTLLAVVAAASQAQVESEAGWGVTDPDLLFAEGTAAAYARGDWPVGLSMER 60  
Db 1 MAVRALKLLTLLAVVAAASQAQVESEAGWGVTDPDLLFAEGTAAAYARGDWPVGLSMER 60

Qy 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAASGAGALRDLSPFGGLLRAACL 120  
Db 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAASGAGALRDLSPFGGLLRAACL 120

Qy 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLQVAYFKINKLEKAVAAAHTEFFVGNPEHMEMQ 180  
Db 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLQVAYFKINKLEKAVAAAHTEFFVGNPEHMEMQ 180

Qy 181 QNLDDYQTMISGVKEADFKDLETQPHMQEFLGLVRLYSEEQPQAVPHLEAALQYFVAYE 240  
Db 181 QNLDDYQTMISGVKEADFKDLETQPHMQEFLGLVRLYSEEQPQAVPHLEAALQYFVAYE 240

Qy 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPFDFL 300  
Db 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPFDFL 300

Qy 301 PSHYNYLOFAYNYGNTQAGECAKTYLLFFPNDEVNQNLAAYAAAMLGEEHTRSGPRE 360  
Db 301 PSHYNYLOFAYNYGNTQAGECAKTYLLFFPNDEVNQNLAAYAAAMLGEEHTRSGPRE 360

Qy 361 SAKEYQRSLLLEKELLFFAYDVFGIPFVDPDSWTPEVIPKRLQEKQKSERETAVRISQE 420  
Db 361 SAKEYQRSLLLEKELLFFAYDVFGIPFVDPDSWTPEVIPKRLQEKQKSERETAVRISQE 420

Qy 421 IGNLMKEITLVEEKTKESLDVSRLEGGPILLYEGISLTMTSKLNGYQRYVMQGVISD 480  
Db 421 IGNLMKEITLVEEKTKESLDVSRLEGGPILLYEGISLTMTSKLNGYQRYVMQGVISD 480

Qy 481 HECQELQRLTNVAATSGDGYRGOTSPHTPNEKFYGVTVFKALKLGOEGKVPLOSAHLYN 540  
Db 481 HECQELQRLTNVAATSGDGYRGOTSPHTPNEKFYGVTVFKALKLGOEGKVPLOSAHLYN 540

Qy 541 VTEKVRIMESYFRDLDTPLFYSHLVCRTAIEEYQAEKDKDSHPVHVNDICILNAETLVC 600  
Db 541 VTEKVRIMESYFRDLDTPLFYSHLVCRTAIEEYQAEKDKDSHPVHVNDICILNAETLVC 600

Qy 601 VKEPPAYTFRDYSAILYLINGDPDGNFYFTELDAKTVTAEVQPCGRAVGSSTENPHG 660  
Db 601 VKEPPAYTFRDYSAILYLINGDPDGNFYFTELDAKTVTAEVQPCGRAVGSSTENPHG 660

Qy 661 VKAVTRGORCAIALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQSOPLDAQOQPPPEP 720  
Db 661 VKAVTRGORCAIALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQSOPLDAQOQPPPEP 720

Qy 721 AQESLSGSESKPKDEL 736  
Db 721 AQESLSGSESKPKDEL 736

RESULT 4  
US-10-045-815-2  
; Sequence 2, Application US/10045815  
; Publication No. US20020160498A1  
; GENERAL INFORMATION:  
; APPLICANT: Wadhwa, Renu  
; APPLICANT: Sugihara, Takashi  
; APPLICANT: Ohide, Akiko  
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
; FILE REFERENCE: 06501-091001  
; CURRENT APPLICATION NUMBER: US/10/045,815  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: PCT/JP00/02731  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: JP 11/118806  
; PRIOR FILING DATE: 1999-04-26  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 363  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-045-815-2

Query Match 48.9%; Score 360; DB 13; Length 363;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAVRALKLLTLLAVVAAASQAQVESEAGWGVTDPDLLFAEGTAAAYARGDWPVGLSMER 60  
Db 1 MAVRALKLLTLLAVVAAASQAQVESEAGWGVTDPDLLFAEGTAAAYARGDWPVGLSMER 60

Qy 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAASGAGALRDLSPFGGLLRAACL 120  
Db 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAASGAGALRDLSPFGGLLRAACL 120

Qy 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLQVAYFKINKLEKAVAAAHTEFFVGNPEHMEMQ 180  
Db 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLQVAYFKINKLEKAVAAAHTEFFVGNPEHMEMQ 180

Qy 181 QNLDDYQTMISGVKEADFKDLETQPHMQEFLGLVRLYSEEQPQAVPHLEAALQYFVAYE 240  
Db 181 QNLDDYQTMISGVKEADFKDLETQPHMQEFLGLVRLYSEEQPQAVPHLEAALQYFVAYE 240

Qy 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPFDFL 300  
Db 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPFDFL 300

Qy 301 PSHYNYLOFAYNYGNTQAGECAKTYLLFFPNDEVNQNLAAYAAAMLGEEHTRSGPRE 360  
Db 301 PSHYNYLOFAYNYGNTQAGECAKTYLLFFPNDEVNQNLAAYAAAMLGEEHTRSGPRE 360

```

RESULT 5
US-10-257-174-25
; Sequence 25, Application US/10257174
; Publication No. US20040034194A1
; GENERAL INFORMATION:
; APPLICANT: Agarwal, Pankaj
; APPLICANT: Murdock, Paul R.
; APPLICANT: Rizvi, Safia K.
; APPLICANT: Smith, Randall F.
; APPLICANT: Xiang, Zhaoxing
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50022
; CURRENT APPLICATION NUMBER: US/10/257,174
; CURRENT FILING DATE: 2002-10-10
; PRIOR APPLICATION NUMBER: PCT/US01/11797
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/196,603
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/199,417
; PRIOR FILING DATE: 2000-04-24
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 25
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-257-174-25

Query Match      37.4%; Score 275; DB 12; Length 708;
Best Local Similarity 99.7%; Pred. No. 5.6e-254;
Matches 375; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 361 SAKYRORSLEKELLFFAYDVFGIPVDPSWTPPEVVKRLOEKQKSERETAVRISQ 420
DB 333 SAKYRORSLEKELLFFAYDVFGIPVDPSWTPPEVVKRLOEKQKSERETAVRISQ 392

QY 421 IGMLMKEIETLVEKTKESLDVSLRTREGGPLLVEGSLTNSKLLNGYQVRVMDGVISD 480
DB 393 IGMLMKEIETLVEKTKESLDVSLRTREGGPLLVEGSLTNSKLLNGSRVMDGVISD 452

QY 481 HECQELQRLTNVAATSGDGRGTQSPHTPNEKFGYVTVFKALXKGQKVPLOSAHLYN 540
DB 453 HECQELQRLTNVAATSGDGRGTQSPHTPNEKFGYVTVFKALXKGQKVPLOSAHLYN 512

QY 541 VTEKVRIMESYFLDTPLYFSYSHLVCRTAIEVQAEKXDDSHPVHVDNCILNAETLVC 600
DB 513 VTEKVRIMESYFLDTPLYFSYSHLVCRTAIEVQAEKXDDSHPVHVDNCILNAETLVC 572

QY 601 VKEPPAYTFRDYSAILYNGDFDGNFYFTELDAKTVTAEVQPCQGRVGFSSGTENPHG 660
DB 573 VKEPPAYTFRDYSAILYNGDFDGNFYFTELDAKTVTAEVQPCQGRVGFSSGTENPHG 632

QY 661 VKAVTRGORCAIALWFTLDPHRSRDRVQADDLVKMLFSPPEMDLSQEQPLDAQCGPPEP 720
DB 633 VKAVTRGORCAIALWFTLDPHRSRDRVQADDLVKMLFSPPEMDLSQEQPLDAQCGPPEP 692

QY 721 AQESLSGSESXPKDEL 736
DB 693 AQESLSGSESXPKDEL 708

RESULT 6
US-09-833-245-1251
; Sequence 1251, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229,358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256,931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199,384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1250
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-245-1250

Query Match      13.7%; Score 101; DB 11; Length 173;
Best Local Similarity 100.0%; Pred. No. 5.9e-88;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAVRALKLLTLLAVVAAASQAEVSEAGWGMVTPDLLFAEGTAAAYARGDWPGVLSMER 60
DB 1 MAVRALKLLTLLAVVAAASQAEVSEAGWGMVTPDLLFAEGTAAAYARGDWPGVLSMER 60

```

```

; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256,931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199,384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1251
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-245-1251

Query Match      29.8%; Score 219; DB 11; Length 359;
Best Local Similarity 99.7%; Pred. No. 1.2e-200;
Matches 319; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAVRALKLLTLLAVVAAASQAEVSEAGWGMVTPDLLFAEGTAAAYARGDWPGVLSMER 60
DB 1 MAVRALKLLTLLAVVAAASQAEVSEAGWGMVTPDLLFAEGTAAAYARGDWPGVLSMER 60

QY 61 ALRSPAALRALRLCRTOCAADFPEWLDPDWSPQAASGAGALRDLSFFGGLLRRACL 120
DB 61 ALRSPAALRALRLCRTOCAADFPEWLDPDWSPQAASGAGALRDLSFFGGLLRRACL 120

QY 121 RRCLGPPAAHSLSEEMELEFRKRSPPYNYQVAYFKINKLEKAVAAATFFVGNPEHMEMQ 180
DB 121 RRCLGPPAAHSLSEEMELEFRKRSPPYNYQVAYFKINKLEKAVAAATFFVGNPEHMEMQ 180

QY 181 QNLDYQTMGKGVKADFKDLETOPHMQEFRGLGVLYSEEQPOEAVPHLEALQEVFVAYE 240
DB 181 QNLDYQTMGKGVKADFKDLETOPHMQEFRGLGVLYSEEQPOEAVPHLEALQEVFVAYE 240

QY 241 ECRALCEGPYDYGNYLYEYNADLFQAITDHYIQLNCKQNCVTELASHPSREKPFDFL 300
DB 241 ECRALCEGPYDYGNYLYEYNADLFQAITDHYIQLNCKQNCVTELASHPSREKPFDFL 300

QY 301 PSHNYLQFAYYNYGNITQA 320
DB 301 PSHNYLQFAYYNYGNITQA 320

RESULT 7
US-09-833-245-1250
; Sequence 1250, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229,358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256,931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199,384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1250
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-245-1250

Query Match      13.7%; Score 101; DB 11; Length 173;
Best Local Similarity 100.0%; Pred. No. 5.9e-88;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAVRALKLLTLLAVVAAASQAEVSEAGWGMVTPDLLFAEGTAAAYARGDWPGVLSMER 60
DB 1 MAVRALKLLTLLAVVAAASQAEVSEAGWGMVTPDLLFAEGTAAAYARGDWPGVLSMER 60

```

QY 61 ALRSRAALRALRLCRCTQCAADFPWELDPDWSFSPAQASGA 101  
|||||  
Db 61 ALRSRAALRALRLCRCTQCAADFPWELDPDWSFSPAQASGA 101  
|||||

RESULT 8  
US-10-045-815-6  
; Sequence 6, Application US/10045815  
; Publication No. US20020160498A1  
; GENERAL INFORMATION:  
; APPLICANT: Wadhwa, Renu  
; APPLICANT: Sugihara, Takashi  
; APPLICANT: Ohide, Akiko  
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
; FILE REFERENCE: 06501-091001  
; CURRENT APPLICATION NUMBER: US/10/045,815  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: PCT/JP00/02731  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: JP 11/118806  
; PRIOR FILING DATE: 1999-04-26  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 747  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-045-815-6

Query Match 12.4%; Score 91; DB 13; Length 747;  
Best Local Similarity 100.0%; Pred. No. 8.2e-78;  
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 612 YSAILYLNGDFDGNFVETLDAKTVTAEVPOCGRAVGFSSGTENPHGVKAVTRGQCA 671  
Db 614 YSAILYLNGDFDGNFYFTLDAKTVTAEVPOCGRAVGFSSGTENPHGVKAVTRGQCA 673  
QY 672 IALWFTLDPHRSRDRVQADDLVKMLFSPEE 702  
Db 674 IALWFTLDPHRSRDRVQADDLVKMLFSPEE 704

RESULT 9  
US-10-045-815-8  
; Sequence 8, Application US/10045815  
; Publication No. US20020160498A1  
; GENERAL INFORMATION:  
; APPLICANT: Wadhwa, Renu  
; APPLICANT: Sugihara, Takashi  
; APPLICANT: Ohide, Akiko  
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
; FILE REFERENCE: 06501-091001  
; CURRENT APPLICATION NUMBER: US/10/045,815  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: PCT/JP00/02731  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: JP 11/118806  
; PRIOR FILING DATE: 1999-04-26  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 542  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-045-815-8

Query Match 11.1%; Score 82; DB 13; Length 542;  
Best Local Similarity 100.0%; Pred. No. 2.5e-69;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 383 FGIPFVDPDSWTPEEVIKRLQEKQSERETAVRISQISGNLMKEITLVEEKTESLDV 442  
|||||

Db 395 FGIPFVDPDSWTPEEVIKRLQEKQSERETAVRISQISGNLMKEITLVEEKTESLDV 444  
QY 443 SRLTREGGGLLYEGISLTMNSK 464  
|||||  
Db 445 SRLTREGGGLLYEGISLTMNSK 466  
|||||

RESULT 10  
US-09-764-875-815  
; Sequence 815, Application US/09764875  
; Publication No. US20040018969A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PJ202  
; CURRENT APPLICATION NUMBER: US/09/764,875  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1249  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 815  
; LENGTH: 219  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-875-815

Query Match 1.8%; Score 13; DB 11; Length 219;  
Best Local Similarity 100.0%; Pred. No. 0.0013;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 602 KEPPAYTFRDYSA 614  
|||||  
Db 140 KEPPAYTFRDYSA 152  
|||||

RESULT 11  
US-09-764-868-736  
; Sequence 736, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT232  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 736  
; LENGTH: 267  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: SITE  
; LOCATION: (6)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-868-736

Query Match 1.8%; Score 13; DB 9; Length 267;  
Best Local Similarity 100.0%; Pred. No. 0.0015;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 602 KEPPAYTFRDYSA 614  
|||||  
Db 154 KEPPAYTFRDYSA 166  
|||||

RESULT 12  
US-09-764-875-1106  
; Sequence 1106, Application US/09764875  
; Publication No. US20040018969A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: P0202  
 ; CURRENT APPLICATION NUMBER: US/09/764,875  
 ; CURRENT FILING DATE: 2001-01-17  
 ; Prior application data removed - consult PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 1249  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 1106  
 ; LENGTH: 267  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (6)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 US-09-764-875-1106

Query Match 1.8%; Score 13; DB 11; Length 267;  
 Best Local Similarity 100.0%; Pred.No. 0.0015;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 602 KEPPAYTRDYSA 614  
 Db 154 KEPPAYTRDYSA 166

## RESULT 13

US-10-012-697-1493  
 ; Sequence 1493, Application US/10012697  
 ; Publication No. US20030215803A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Escobedo, Jaime  
 ; APPLICANT: Garcia, Pablo Dominguez  
 ; APPLICANT: Kassam, Altaf  
 ; APPLICANT: Lamson, George  
 ; APPLICANT: Scott, Beth  
 ; APPLICANT: Drmanac, Radoje  
 ; APPLICANT: Crkvenjakov, Radomir  
 ; APPLICANT: Dickson, Mark  
 ; APPLICANT: Drmanac, Snezana  
 ; APPLICANT: Labat, Ivan  
 ; APPLICANT: Leshkowitz, Dena  
 ; APPLICANT: Kita, David  
 ; APPLICANT: Garcia, Veronica  
 ; APPLICANT: Jones, Lee William  
 ; APPLICANT: Stache-Crain, Birgit  
 ; TITLE OF INVENTION: HUMAN GENES AND GENE EXPRESSION PRODUCTS  
 ; FILE REFERENCE: 2300-16252  
 ; CURRENT APPLICATION NUMBER: US/10/012,697  
 ; CURRENT FILING DATE: 2003-01-21  
 ; PRIOR APPLICATION NUMBER: 60/254,648  
 ; PRIOR FILING DATE: 2000-12-07  
 ; PRIOR APPLICATION NUMBER: 60/275,668  
 ; PRIOR FILING DATE: 2001-03-13  
 ; NUMBER OF SEQ ID NOS: 1568  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1493  
 ; LENGTH: 282  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-012-697-1493

Query Match 1.8%; Score 13; DB 15; Length 282;  
 Best Local Similarity 100.0%; Pred.No. 0.0016;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 602 KEPPAYTRDYSA 614  
 Db 169 KEPPAYTRDYSA 181

## RESULT 14

US-10-094-749-2560  
 ; Sequence 2560, Application US/10094749  
 ; Publication No. US20030219741A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ISOGAI, TAKAO  
 ; APPLICANT: SUGIYAMA, TOMOYASU  
 ; APPLICANT: OTSUKI, TETSUJI  
 ; APPLICANT: WAKAMATSU, AI  
 ; APPLICANT: SATO, HIROYUKI  
 ; APPLICANT: ISHII, SHIZUKO  
 ; APPLICANT: YAMAMOTO, JUN-ICHI  
 ; APPLICANT: ISONO, YUUKO  
 ; APPLICANT: HIO, YURI  
 ; APPLICANT: OTSUKA, KAORU  
 ; APPLICANT: NAGAI, KEIICHI  
 ; APPLICANT: IRIE, RYOTARO  
 ; APPLICANT: TAMECHIKA, ICHIRO  
 ; APPLICANT: SEKI, NAOHICO  
 ; APPLICANT: YOSHIKAWA, TSUTOMU  
 ; APPLICANT: OTSUKA, MOTOKYUKI  
 ; APPLICANT: NAGAHARI, KENJI  
 ; APPLICANT: MASUHO, YASUHIKO  
 ; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
 ; FILE REFERENCE: 084335/0160  
 ; CURRENT APPLICATION NUMBER: US/10/094,749  
 ; CURRENT FILING DATE: 2002-03-12  
 ; PRIOR APPLICATION NUMBER: 60/350,435  
 ; PRIOR FILING DATE: 2002-01-24  
 ; PRIOR APPLICATION NUMBER: JP 2001-328381  
 ; PRIOR FILING DATE: 2001-09-14  
 ; NUMBER OF SEQ ID NOS: 3381  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 2560  
 ; LENGTH: 708  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-094-749-2560

Query Match 1.8%; Score 13; DB 15; Length 708;  
 Best Local Similarity 100.0%; Pred.No. 0.0036;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 602 KEPPAYTRDYSA 614  
 Db 595 KEPPAYTRDYSA 607

## RESULT 15

US-09-572-404B-1740  
 ; Sequence 1740, Application US/09572404B  
 ; Publication No. US20030078374A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Proteom Ltd  
 ; TITLE OF INVENTION: Complementary peptide ligands from the human genome  
 ; FILE REFERENCE: Human patent  
 ; CURRENT APPLICATION NUMBER: US/09/572,404B  
 ; CURRENT FILING DATE: 2000-05-17  
 ; NUMBER OF SEQ ID NOS: 4203  
 ; SOFTWARE: ProtPatent version 1.0  
 ; SEQ ID NO 1740  
 ; LENGTH: 10  
 ; TYPE: PRT  
 ; ORGANISM: Homo Sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: sequence located in NOP56 at 110-119 and may interact with Sequenc  
 ; OTHER INFORMATION: in this patent.  
 US-09-572-404B-1740

Query Match 1.1%; Score 8; DB 10; Length 10;  
 Best Local Similarity 100.0%; Pred.No. 4.9;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 116 RAACLRRRC 123

Db           |||||||  
              3 RAAOLREC 10

Search completed: July 18, 2004, 10:02:04  
Job time : 54 secs

This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 09:41:18 ; Search time 685 Seconds  
(without alignments)  
335.832 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 3870  
Sequence: 1 MAVRALKLTLLAVVAAAS.....PPEPAQESLSGSESKPKDEL 736

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1285345 seqs, 312560633 residues

Total number of hits satisfying chosen parameters: 1285345

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Published Applications AA:\*
- 1: /cgn2\_6/prodata/2/pubpaa/US07\_PUBCOMB.pep.\*
  - 2: /cgn2\_6/prodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/prodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
  - 4: /cgn2\_6/prodata/2/pubpaa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/prodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/prodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/prodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
  - 8: /cgn2\_6/prodata/2/pubpaa/US08\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/prodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
  - 10: /cgn2\_6/prodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/prodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
  - 12: /cgn2\_6/prodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
  - 13: /cgn2\_6/prodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
  - 14: /cgn2\_6/prodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
  - 15: /cgn2\_6/prodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
  - 16: /cgn2\_6/prodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
  - 17: /cgn2\_6/prodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
  - 18: /cgn2\_6/prodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3870	100.0	736	13	US-10-045-815-4
2	3846	99.4	736	12	US-10-257-174-26
3	3846	99.4	736	12	US-10-312-352-33
4	3572	92.3	708	12	US-10-257-174-25
5	3316	85.7	747	13	US-10-045-815-6
6	2384	61.6	542	13	US-10-045-815-8
7	1911	49.4	363	13	US-10-045-815-2
8	1800	46.5	359	11	US-09-833-245-1251
9	1751	45.2	708	15	US-10-094-749-2560
10	1473	38.1	736	15	US-10-094-886-24
11	906.5	23.4	282	15	US-10-012-697-1493
12	836.5	21.6	267	9	US-09-764-868-736
13	836.5	21.6	267	11	US-09-764-875-1106
14	657	17.0	173	11	US-09-833-245-1250
15	612	15.8	219	11	US-09-764-875-815

16	567	14.7	401	9	US-09-729-674-2	Sequence 2, Appli
17	501	12.9	528	15	US-10-094-749-2363	Sequence 2363, Ap
18	383.5	9.9	153	12	US-10-276-774-1567	Sequence 1567, Ap
19	274	7.1	262	15	US-10-108-260A-4501	Sequence 4501, Ap
20	183.5	4.7	104	15	US-10-108-260A-3928	Sequence 3928, Ap
21	177	4.6	241	12	US-10-424-599-224808	Sequence 224808,
22	148.5	3.8	603	15	US-10-289-762-1011	Sequence 1011, Ap
23	135	3.5	1207	12	US-10-250-823-8	Sequence 8, Appli
24	135	3.5	1207	12	US-10-250-823-10	Sequence 10, Appli
25	129	3.3	227	9	US-09-782-980-79	Sequence 79, Appli
26	126.5	3.3	1154	12	US-10-250-823-4	Sequence 4, Appli
27	126.5	3.3	1154	12	US-10-250-823-6	Sequence 6, Appli
28	123	3.2	2771	16	US-10-437-963-195406	Sequence 195406,
29	122	3.2	1529	16	US-10-389-566-2103	Sequence 2103, Ap
30	119.5	3.1	1905	15	US-10-259-194A-86	Sequence 86, Appl
31	118.5	3.1	847	12	US-10-262-511-8	Sequence 8, Appli
32	118.5	3.1	2053	13	US-10-017-216-2	Sequence 2, Appli
33	118.5	3.1	2053	14	US-10-325-430-12	Sequence 12, Appl
34	118	3.0	1441	15	US-10-412-897-3	Sequence 3, Appli
35	118	3.0	1597	12	US-09-364-956-41	Sequence 41, Appl
36	118	3.0	1597	13	US-10-017-216-6	Sequence 6, Appli
37	118	3.0	1641	12	US-09-964-956-40	Sequence 40, Appl
38	118	3.0	1641	13	US-10-017-216-5	Sequence 5, Appli
39	118	3.0	2055	13	US-10-017-216-4	Sequence 4, Appli
40	117	3.0	832	12	US-10-262-511-6	Sequence 6, Appli
41	117	3.0	1286	12	US-09-964-956-38	Sequence 38, Appl
42	117	3.0	1286	13	US-10-017-216-7	Sequence 7, Appli
43	117	3.0	1958	13	US-10-028-946-4	Sequence 4, Appli
44	117	3.0	2053	12	US-09-964-956-11	Sequence 11, Appl
45	117	3.0	2053	12	US-10-262-511-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-10-045-815-4  
; Sequence 4, Application US/10045815  
; Publication No. US20020160498A1  
; GENERAL INFORMATION:  
; APPLICANT: Wadwa, Renu  
; APPLICANT: Sugihara, Takashi  
; APPLICANT: Ohide, Akiko  
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
; FILE REFERENCE: 06501-091001  
; CURRENT APPLICATION NUMBER: US/10/045,815  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: PCT/JF00/02731  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: JP 11/118806  
; PRIOR FILING DATE: 1999-04-26  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 736  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-045-815-4

Query Match	100.0%	Score	3870;	DB	13;	Length	736;
Best Local Similarity	100.0%;	Pred. No.	0;				
Matches	736;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	MAVRALKLTLLAVVAAASQAESVSEAGWGMVTPDLLFAEGTAAAYARGDWPGVLSMER	60				
Db	1	MAVRALKLTLLAVVAAASQAESVSEAGWGMVTPDLLFAEGTAAAYARGDWPGVLSMER	60				
QY	61	ALRSPAALRALRLCRTCAADFPWELDPWSPQAQSGAGALRDLSPFGGLLRAACL	120				
Db	61	ALRSPAALRALRLCRTCAADFPWELDPWSPQAQSGAGALRDLSPFGGLLRAACL	120				
QY	121	RRCLGPPAAHSLSEMELEFRKSPYNYQVAYFKINKLEKXAAAAHTFFVGNPEHMEMQ	180				

```
Db 121 RRLGPPAAHSLSEMELEFRKRSPPYNYLVAVFKINKLEKAVAAAHTFFVGNPEHMEMQ 180
Qy 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGRLVLYSEEPQAVPHLEAALQEFVAYE 240
Db 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGRLVLYSEEPQAVPHLEAALQEFVAYE 240
Qy 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTELASHPRSRKPPEDEL 300
Db 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTELASHPRSRKPPEDEL 300
Qy 301 PSHNYLQFAYYNYGNTQAGECAKTYLLFPNDENVNQLAYYAAMLGEHTRSIGPRE 360
Db 301 PSHNYLQFAYYNYGNTQAGECAKTYLLFPNDENVNQLAYYAAMLGEHTRSIGPRE 360
Qy 361 SAKYRQSRSLLEKELLFPAYDVFGIPFVDPDSWTPPEEVIKRLQEKOKSERETAVRISQ 420
Db 361 SAKYRQSRSLLEKELLFPAYDVFGIPFVDPDSWTPPEEVIKRLQEKOKSERETAVRISQ 420
Qy 421 IGMLMKEIETLVEBKTESLDVSRLTREGGPLLVEGSLTMNSKLLNGYQVVMGVDGVID 480
Db 421 IGMLMKEIETLVEBKTESLDVSRLTREGGPLLVEGSLTMNSKLLNGYQVVMGVDGVID 480
Qy 481 HECQELQRLTNVAATSGDGYRGQTSPTPNEKFGYVTVFKALKLGQBGKVPLOSAHLYN 540
Db 481 HECQELQRLTNVAATSGDGYRGQTSPTPNEKFGYVTVFKALKLGQBGKVPLOSAHLYN 540
Qy 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIBEVOAERKDDSHPHVHVDNCILNAETLVC 600
Db 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIBEVOAERKDDSHPHVHVDNCILNAETLVC 600
Qy 601 VKEPPATFRDYSAILYLNGDFDGNFYFTELDKATVTAEVQPCGQRAVGFSSGTENPHG 660
Db 601 VKEPPATFRDYSAILYLNGDFDGNFYFTELDKATVTAEVQPCGQRAVGFSSGTENPHG 660
Qy 661 VKAVTRGORCAIALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEPDLAQOQGPPEP 720
Db 661 VKAVTRGORCAIALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEPDLAQOQGPPEP 720
Qy 721 AQESLSGSESXPKDEL 736
Db 721 AQESLSGSESXPKDEL 736
```

## RESULT 2

```
US-10-257-174-26
; Sequence 26, Application US/10257174
; Publication No. US20040034194A1
; GENERAL INFORMATION:
; APPLICANT: Agarwal, Pankaj
; APPLICANT: Murdock, Paul R.
; APPLICANT: Rizvi, Safia K.
; APPLICANT: Smith, Randall F.
; APPLICANT: Xiang, Zhaoying
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50022
; CURRENT APPLICATION NUMBER: US/10/257,174
; CURRENT FILING DATE: 2002-10-10
; PRIOR APPLICATION NUMBER: PCT/US01/11797
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/196,603
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/199,417
; PRIOR FILING DATE: 2000-04-24
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Fast-SEQ for Windows Version 3.0
; SEQ ID NO 26
; LENGTH: 736
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
Query Match 99.4%; Score 3846; DB 12; Length 736;
Best Local Similarity 99.6%; Pred. No. 0;
```

```
Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 MAVRALKLLTLLAVAAAASQAEVESEAGWGTPTDLLFAEGTAAYARGDWPVGLSNMER 60
Db 1 MAVRALKLLTLLAVAAAASQAEVESEAGWGTPTDLLFAEGTAAYARGDWPVGLSNMER 60
Qy 61 ALRSPAAALRALRLCRTOCAADFFWELDPDWSPPAQASGAGALRDLSPFGGLRRAACL 120
Db 61 ALRSPAAALRALRLCRTOCAADFFWELDPDWSPPAQASGAGALRDLSPFGGLRRAACL 120
Qy 121 RRLGPPAAHSLSEMELEFRKRSPPYNYLVAVFKINKLEKAVAAAHTFFVGNPEHMEMQ 180
Db 121 RRLGPPAAHSLSEMELEFRKRSPPYNYLVAVFKINKLEKAVAAAHTFFVGNPEHMEMQ 180
Qy 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGRLVLYSEEPQAVPHLEAALQEFVAYE 240
Db 181 QNLDDYQTMGSKVKEADFKDLETPHMQEFLGRLVLYSEEPQAVPHLEAALQEFVAYE 240
Qy 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTELASHPRSRKPPEDEL 300
Db 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTELASHPRSRKPPEDEL 300
Qy 301 PSHNYLQFAYYNYGNTQAGECAKTYLLFPNDENVNQLAYYAAMLGEHTRSIGPRE 360
Db 301 PSHNYLQFAYYNYGNTQAGECAKTYLLFPNDENVNQLAYYAAMLGEHTRSIGPRE 360
Qy 361 SAKYRQSRSLLEKELLFPAYDVFGIPFVDPDSWTPPEEVIKRLQEKOKSERETAVRISQ 420
Db 361 SAKYRQSRSLLEKELLFPAYDVFGIPFVDPDSWTPPEEVIKRLQEKOKSERETAVRISQ 420
Qy 421 IGMLMKEIETLVEBKTESLDVSRLTREGGPLLVEGSLTMNSKLLNGYQVVMGVDGVID 480
Db 421 IGMLMKEIETLVEBKTESLDVSRLTREGGPLLVEGSLTMNSKLLNGYQVVMGVDGVID 480
Qy 481 HECQELQRLTNVAATSGDGYRGQTSPTPNEKFGYVTVFKALKLGQBGKVPLOSAHLYN 540
Db 481 HECQELQRLTNVAATSGDGYRGQTSPTPNEKFGYVTVFKALKLGQBGKVPLOSAHLYN 540
Qy 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIBEVOAERKDDSHPHVHVDNCILNAETLVC 600
Db 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIBEVOAERKDDSHPHVHVDNCILNAETLVC 600
Qy 601 VKEPPATFRDYSAILYLNGDFDGNFYFTELDKATVTAEVQPCGQRAVGFSSGTENPHG 660
Db 601 VKEPPATFRDYSAILYLNGDFDGNFYFTELDKATVTAEVQPCGQRAVGFSSGTENPHG 660
Qy 661 VKAVTRGORCAIALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEPDLAQOQGPPEP 720
Db 661 VKAVTRGORCAIALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEPDLAQOQGPPEP 720
Qy 721 AQESLSGSESXPKDEL 736
Db 721 AQESLSGSESXPKDEL 736
```

## RESULT 3

```
US-10-312-352-33
; Sequence 33, Application US/10312352
; Publication No. US20040053824A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.; TANG, Y. Tom
; APPLICANT: YUE, Henry; AZIMZAI, Yalda
; APPLICANT: HE, Ann; BATRA, Sajeev
; APPLICANT: LO, Terence P.; NGUYEN, Dannel B.
; APPLICANT: BURRILL, John D.; MARCUS, Gregory A.
; APPLICANT: ZINGLER, Kurt A.; GANDHI, Ameena R.
; APPLICANT: LAL, Preeti G.; KEARNEY, Liam
; APPLICANT: BURFORD, Neil; YAO, Monique G.
; APPLICANT: CHAWLA, Narinder K.; ELLIOT, Vicki S.
; APPLICANT: ARVIZU, Chandra S.; KHAN, Farrah A.
; APPLICANT: BAUGHN, Mariah R.; HAFALIA, April, J.A.
; APPLICANT: POITICKY, Jennifer L.; AU-YOUNG, Janice K.
; APPLICANT: LU, Yan; BOROWSKY, Mark L.
```

APPLICANT: LU, Dyung Aina M.; RAMKUMAR, Jayalaxmi  
APPLICANT: YANG, Junming; GURURAJAN, Rajagopal  
APPLICANT: WARREN, Bridget A.; GIETZEN, Kimberly J.  
APPLICANT: XU, Yuming; KALLICK, Deborah A.  
APPLICANT: LEE, Ernestine A.; THANGAVELU, Kavitha  
APPLICANT: DELEGANE, Angelo M.; LEE, Sally  
TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES  
FILE REFERENCE: PF-0794 USN  
CURRENT APPLICATION NUMBER: US/10/312,352  
CURRENT FILING DATE: 2002-12-18  
PRIORITY APPLICATION NUMBER: PCT/US01/21067  
PRIORITY FILING DATE: 2001-06-29  
PRIORITY APPLICATION NUMBER: US 60/215,454  
PRIORITY FILING DATE: 2000-06-30  
PRIORITY APPLICATION NUMBER: US 60/219,462  
PRIORITY FILING DATE: 2000-07-18  
PRIORITY APPLICATION NUMBER: US 60/240,111  
PRIORITY FILING DATE: 2000-10-12  
PRIORITY APPLICATION NUMBER: US 60/240,106  
PRIORITY FILING DATE: 2000-10-12  
PRIORITY APPLICATION NUMBER: US 60/244,021  
PRIORITY FILING DATE: 2000-10-27  
PRIORITY APPLICATION NUMBER: US 60/248,887  
PRIORITY FILING DATE: 2000-11-14  
PRIORITY APPLICATION NUMBER: US 60/249,570  
PRIORITY FILING DATE: 2000-11-16  
NUMBER OF SEQ ID NOS: 72  
SOFTWARE: PERL Program  
SEQ ID NO 33  
LENGTH: 736  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
OTHER INFORMATION: Incyte ID No. US20040053824A1 6780147CD1  
US-10-312-352-33

Query Match 99.4%; Score 3846; DB 12; Length 736;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAVRALKLLTLLAVVAAASQAQAEVESEAGWMTPTDLLFAEGTAAAYARGDWPGVLSMER 60  
DB 1 MAVRALKLLTLLAVVAAASQAQAEVESEAGWMTPTDLLFAEGTAAAYARGDWPGVLSMER 60

QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
DB 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120

QY 121 RRLCLGPPAAHLSSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHTEFFVGNPEHMEQ 180  
DB 121 RRLCLGPPAAHLSSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHTEFFVGNPEHMEQ 180

QY 181 QNLDYQYQTMGSVKGEADFKDLETPHMQEFLGLVRLYSESQPOQAVPHLEAALQEVYVAYE 240  
DB 181 QNLDYQYQTMGSVKGEADFKDLETPHMQEFLGLVRLYSESQPOQAVPHLEAALQEVYVAYE 240

QY 241 ECRALCEGPDYDGYNYLYNADLFQAITDHYIQVNLCKQNCVTELASHPSREKPFDFL 300  
DB 241 ECRALCEGPDYDGYNYLYNADLFQAITDHYIQVNLCKQNCVTELASHPSREKPFDFL 300

QY 301 PSHYNYLOFAYNYIGNYTOAGECAKTYLLFPNDVNMQNLAYYAAAMLGSEHTRTSIGPRE 360  
DB 301 PSHYNYLOFAYNYIGNYTOAGECAKTYLLFPNDVNMQNLAYYAAAMLGSEHTRTSIGPRE 360

QY 361 SAKYEQRSLLLEKELLFFAYDVDFGIPFVDSWTPTEVTPKRLQEKQKSERETAVRISQE 420  
DB 361 SAKYEQRSLLLEKELLFFAYDVDFGIPFVDSWTPTEVTPKRLQEKQKSERETAVRISQE 420

QY 421 IGNLMKEIETLVEKTKESLDVSRLTREGGPLLVEGILSTMTSKLLNGVQRVMDGVISD 480  
DB 421 IGNLMKEIETLVEKTKESLDVSRLTREGGPLLVEGILSTMTSKLLNGVQRVMDGVISD 480

QY 481 HECQELQRLTNVAATSGDGYRGQTSPTHTPNKFKYGVTVFKALKLQBGKVPLOSAHLYYN 540  
DB 481 HECQELQRLTNVAATSGDGYRGQTSPTHTPNKFKYGVTVFKALKLQBGKVPLOSAHLYYN 540

QY 541 VTEKVRRTMESYFRILDTPLYESYSHLVCRTAIEVQAEKDDSHPVHVDNCLNAETLVC 600  
DB 541 VTEKVRRTMESYFRILDTPLYESYSHLVCRTAIEVQAEKDDSHPVHVDNCLNAETLVC 600

QY 601 VKEPPAYTFRDYSAILYLNGDFDGGNFYFTELDKATVTAEVQPCGRAVGFSSGTENPHG 660  
DB 601 VKEPPAYTFRDYSAILYLNGDFDGGNFYFTELDKATVTAEVQPCGRAVGFSSGTENPHG 660

QY 661 VKAVTRGORCAIALWFTLDPRHSERDRVOADDIVKMLFSPPEMDLSQEOPLDAQGGPPEP 720  
DB 661 VKAVTRGORCAIALWFTLDPRHSERDRVOADDIVKMLFSPPEMDLSQEOPLDAQGGPPEP 720

QY 721 AQESLSGSESKEPKDEL 736  
DB 721 AQESLSGSESKEPKDEL 736

RESULT 4  
US-10-257-174-25  
Sequence 25, Application US/10257174  
Publication No. US20040034194A1  
GENERAL INFORMATION:  
APPLICANT: Agarwal, Pankaj  
APPLICANT: Murdoch, Paul R.  
APPLICANT: Rizvi, Safia K.  
APPLICANT: Smith, Randall F.  
APPLICANT: Xiang, Zhaoying  
TITLE OF INVENTION: NOVEL COMPOUNDS  
FILE REFERENCE: GPS0022  
CURRENT APPLICATION NUMBER: US/10/257,174  
CURRENT FILING DATE: 2002-10-10  
PRIORITY APPLICATION NUMBER: PCT/US01/11797  
PRIORITY FILING DATE: 2001-04-11  
PRIORITY APPLICATION NUMBER: 60/196,603  
PRIORITY FILING DATE: 2000-04-13  
PRIORITY APPLICATION NUMBER: 60/199,417  
PRIORITY FILING DATE: 2000-04-24  
NUMBER OF SEQ ID NOS: 48  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 25  
LENGTH: 708  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-257-174-25

Query Match 92.3%; Score 3572; DB 12; Length 708;  
Best Local Similarity 93.9%; Pred. No. 0;  
Matches 691; Conservative 5; Mismatches 12; Indels 28; Gaps 3;

QY 1 MAVRALKLLTLLAVVAAASQAQAEVESEAGWMTPTDLLFAEGTAAAYARGDWPGVLSMER 60  
DB 1 MAVRALKLLTLLAVVAAASQAQAEVESEAGWMTPTDLLFAEGTAAAYARGDWPGVLSMER 60

QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
DB 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120

QY 121 RRLCLGPPAAHLSSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHTEFFVGNPEHMEQ 180  
DB 121 RRLCLGPPAAHLSSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHTEFFVGNPEHMEQ 180

QY 181 QNLDYQYQTMGSVKGEADFKDLETPHMQEFLGLVRLYSESQPOQAVPHLEAALQEVYVAYE 240  
DB 181 QNLDYQYQTMGSVKGEADFKDLETPHMQEFLGLVRLYSESQPOQAVPHLEAALQEVYVAYE 240

QY 241 ECRALCEGPDYDGYNYLYNADLFQAITDHYIQVNLCKQNCVTELASHPSREKPFDFL 300  
DB 241 ECRALCEGPDYDGYNYLYNADLFQAITDHYIQVNLCKQNCVTELASHPSREKPFDFL 300

```
QY 301 PSHNYLQFAYNNTGNVTOGECACKTYLLPFPNDENVNQNLAAYAAAMLGEEHTRSTIGPRE 360
DB 301 PSHNYLQFAYNNTGNVTOGECACKTYLLPFPNDENVNQNLAAYAAAMLGEEHTRSTIGPRE 360
QY 361 SAKYRORSLEKELLFPAYDVFGIPFVDPDSWTPPEEVIKRLQEKOKSRETAVALRSQE 420
DB 361 SAKYRORSLEKELLFPAYDVFGIPFVDPDSWTPPEEVIKRLQEKOKSRETAVALRSQE 420
QY 333 SAKYRORSLEKELLFPAYDVFGIPFVDPDSWTPPEEVIKRLQEKOKSRETAVALRSQE 332
DB 333 SAKYRORSLEKELLFPAYDVFGIPFVDPDSWTPPEEVIKRLQEKOKSRETAVALRSQE 332
QY 421 IGLNLMKEIETLVEEKTESLDVSLTREGGPLYEGISLTWNSKLLNGYQVVMVDGVID 480
DB 421 IGLNLMKEIETLVEEKTESLDVSLTREGGPLYEGISLTWNSKLLNGYQVVMVDGVID 480
QY 481 HECQELQRLTNVAATSGDYGRTSPHTPNEKFGVTVFKALKLQEGKVPLOSALHYLN 540
DB 481 HECQELQRLTNVAATSGDYGRTSPHTPNEKFGVTVFKALKLQEGKVPLOSALHYLN 540
QY 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEBEOAERKDDSHPVHVDNCILNAETLVC 600
DB 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEBEOAERKDDSHPVHVDNCILNAETLVC 600
QY 601 VKEPPATFRDYSAILYLNGDFDGCNFFYFTLDAKTVAEVOQCGRAVGFSSGTENPHG 660
DB 601 VKEPPATFRDYSAILYLNGDFDGCNFFYFTLDAKTVAEVOQCGRAVGFSSGTENPHG 660
QY 573 VKEPPATFRDYSAILYLNGDFDGCNFFYFTLDAKTVAEVOQCGRAVGFSSGTENPHG 632
DB 573 VKEPPATFRDYSAILYLNGDFDGCNFFYFTLDAKTVAEVOQCGRAVGFSSGTENPHG 632
QY 661 VKAVTRGQRCALALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSOEQPLDAAQCGPP 720
DB 661 VKAVTRGQRCALALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSOEQPLDAAQCGPP 720
QY 721 AQESLSGSESKPKDEL 736
DB 721 AQESLSGSESKPKDEL 736
QY 693 AQESLSGSESKPKDEL 708
DB 693 AQESLSGSESKPKDEL 708

RESULT 5
US-10-045-815-6
; Sequence 6, Application US/10045815
; Publication No. US20020160498A1
; GENERAL INFORMATION:
; APPLICANT: Wadhwa, Renu
; APPLICANT: Sugihara, Takashi
; APPLICANT: Ohide, Akiko
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; FILE REFERENCE: 06501-091001
; CURRENT APPLICATION NUMBER: US/10/045,815
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: PCT/JP00/02731
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: JP 11/118806
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 747
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-045-815-6

Query Match 85.7%; Score 3316; DB 13; Length 747;
Best Local Similarity 88.2%; Pred. No. 2.9e-296;
Matches 623; Conservative 33; Mismatches 48; Indels 2; Gaps 1;

QY 25 ESEAGWGMVTPDLLFAEGTAAYARGDWPVGLSMERALSRAALRLRCRTQCAADFP 84
DB 25 ESEAGWGMVTPDLLFAEGTAAYARGDWPVGLSMERALSRAALRLRCRTQCAADFP 84
QY 85 WELDPDWSPP--AQASGAGALRDLSPFGGLLRRACLRCLGPPAAHSLSEMELEFRK 142
DB 85 WELDPDWSPP--AQASGAGALRDLSPFGGLLRRACLRCLGPPAAHSLSEMELEFRK 142
QY 143 RSPNYLQVAYFKINKLEKAVAAAHFFVGNPHMEMQNLDDYQTMGSKVKEADFKDLET 202
DB 143 RSPNYLQVAYFKINKLEKAVAAAHFFVGNPHMEMQNLDDYQTMGSKVKEADFKDLET 202
QY 145 RSPNYLQVAYFKINKLEKAVAAAHFFVGNPHMEMQNLDDYQTMGSKVKEADFKDLET 204
DB 145 RSPNYLQVAYFKINKLEKAVAAAHFFVGNPHMEMQNLDDYQTMGSKVKEADFKDLET 204
```

```
QY 203 QPHMQEFLGRLVLYSEBQEQEAVPHLEAALQBYFVAYVEECALCEGPPYDYGNYLEYNA 262
DB 203 QPHMQEFLGRLVLYSEBQEQEAVPHLEAALQBYFVAYVEECALCEGPPYDYGNYLEYNA 262
QY 263 DLFOAITDHYIQVNLCKNCVTELASHESREKPEDFLPSHYNVLOFAYNIGNVTOAGE 322
DB 263 DLFOAITDHYIQVNLCKNCVTELASHESREKPEDFLPSHYNVLOFAYNIGNVTOAGE 322
QY 323 CAKTYLLFPFNDENVNQNLAAYAAAMLGEEHTRSTIGPRESAKYRORSLEKELLFPAYDV 382
DB 323 CAKTYLLFPFNDENVNQNLAAYAAAMLGEEHTRSTIGPRESAKYRORSLEKELLFPAYDV 382
QY 383 FGIPEVDSDSWTPPEEVIKRLQEKOKSRETAVALRSQEGNLMKEIETLVEEKTESLDV 442
DB 383 FGIPEVDSDSWTPPEEVIKRLQEKOKSRETAVALRSQEGNLMKEIETLVEEKTESLDV 442
QY 443 SRLTREGGPLYEGISLTWNSKLLNGYQVVMVDGVIDSDCEQELQRLTNVAATSGDYG 502
DB 443 SRLTREGGPLYEGISLTWNSKLLNGYQVVMVDGVIDSDCEQELQRLTNVAATSGDYG 502
QY 503 QTSPTPNEKFGVTVFKALKLQEGKVPLOSALHYLNVTETKVRIMESYFRDLTPLYFS 562
DB 503 QTSPTPNEKFGVTVFKALKLQEGKVPLOSALHYLNVTETKVRIMESYFRDLTPLYFS 562
QY 563 YSHLVCRTAIEBEOAERKDDSHPVHVDNCILNAETLVCVKEPPATFRDYSAILYLNGDF 622
DB 563 YSHLVCRTAIEBEOAERKDDSHPVHVDNCILNAETLVCVKEPPATFRDYSAILYLNGDF 622
QY 623 DGCNFFYFTLDAKTVAEVOQCGRAVGFSSGTENPHGKAVTRGQRCALALWFTLDPH 682
DB 623 DGCNFFYFTLDAKTVAEVOQCGRAVGFSSGTENPHGKAVTRGQRCALALWFTLDPH 682
QY 683 SERDRVQADDLVKMLFSPPEMDLSOEQPLDAAQCGPPPAQESLSGS 728
DB 683 SERDRVQADDLVKMLFSPPEMDLSOEQPLDAAQCGPPPAQESLSGS 728

RESULT 6
US-10-045-815-8
; Sequence 8, Application US/10045815
; Publication No. US20020160498A1
; GENERAL INFORMATION:
; APPLICANT: Wadhwa, Renu
; APPLICANT: Sugihara, Takashi
; APPLICANT: Ohide, Akiko
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; FILE REFERENCE: 06501-091001
; CURRENT APPLICATION NUMBER: US/10/045,815
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: PCT/JP00/02731
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: JP 11/118806
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 542
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-045-815-8

Query Match 61.6%; Score 2384; DB 13; Length 542;
Best Local Similarity 87.9%; Pred. No. 1.7e-210;
Matches 451; Conservative 25; Mismatches 35; Indels 2; Gaps 1;

QY 25 ESEAGWGMVTPDLLFAEGTAAYARGDWPVGLSMERALSRAALRLRCRTQCAADFP 84
DB 25 ESEAGWGMVTPDLLFAEGTAAYARGDWPVGLSMERALSRAALRLRCRTQCAADFP 84
QY 85 WELDPDWSPP--AQASGAGALRDLSPFGGLLRRACLRCLGPPAAHSLSEMELEFRK 142
DB 85 WELDPDWSPP--AQASGAGALRDLSPFGGLLRRACLRCLGPPAAHSLSEMELEFRK 142
QY 145 RSPNYLQVAYFKINKLEKAVAAAHFFVGNPHMEMQNLDDYQTMGSKVKEADFKDLET 202
DB 145 RSPNYLQVAYFKINKLEKAVAAAHFFVGNPHMEMQNLDDYQTMGSKVKEADFKDLET 202
```

QY 143 RSPYNYLOVAYFKINKLEKAVAAAHFTFVGNBEHMEMOONLDYYOTMSGVKEADFKDLET 202  
DB 145 RSPYNYLOVAYFKINKLEKAVAAAHFTFVGNBEHMEMOONLDYYOTMSGVKEADFKDLEA 204  
QY 203 QPHMOEFLRGVLYSEBEOQEAAPHLEAALQYFVAYBECRALCEGPDYDYGYNLYEYNA 262  
DB 205 KPHMHEFLRGVLYSEBEOQEAAPHLEAALQYFVAYBECRALCEGPDYDYGYNLYEYNA 264  
QY 263 DLFOAITDHYIOVLNCKQNCVTELASHPSREKPFDFLPDLSHYNLYOFAYYNGNTOAGE 322  
DB 265 DLFOAITDHYIOVLNCKQNCVTELASHPSREKPFDFLPDLSHYNLYOFAYYNGNTOAGE 324  
QY 323 CAKTYLLFPNDEVNMNOLAYVAALGEEHTRISIGPRESAKERYORSLEKELLFFAYDV 382  
DB 325 CAKTYLLFPNDEVNMNOLAYVAALGEEHTRISIGPRESAKERYORSLEKELLFFAYDI 384  
QY 383 FGIPFVDPDSWTEPEVPIKLOEKOKSRETAVRISQETGNLMKEIETLVEKTKWESLDV 442  
DB 385 FGIPFVDPDSWTEPEVPIKLOEKOKSRETAVRISQETGNLMKEIETLVEKTKWESLDV 444  
QY 443 SRLTREGGPLYVIGISLTWNSKLINGYQVWMDGVISDHECOELORLTNVAATSGDGYRG 502  
DB 445 SRLTREGGPLYVIGISLTWNSKLINGYQVWMDGVISDHECOELORLTNVAATSGDGYRG 504  
QY 503 QTSPTPNEKFGYGVTVFKALKGQEGKVPLOSA 535  
DB 505 QTSPTPNEKFGYGVTVFKALKGQEGKVPLOSA 537

RESULT 7  
US-10-045-815-2  
; Sequence 2, Application US/10045815  
; Publication No. US20020160498A1  
; GENERAL INFORMATION:  
; APPLICANT: Wadhwa, Renu  
; APPLICANT: Sugihara, Takashi  
; APPLICANT: Ohide, Akiko  
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
; FILE REFERENCE: 06501-091001  
; CURRENT APPLICATION NUMBER: US/10/045,815  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: PCT/JP00/02731  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: JP 11/118806  
; PRIOR FILING DATE: 1999-04-26  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Fast-SEQ for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 363  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-045-815-2

Query Match 49.4%; Score 1911; DB 13; Length 363;  
Best Local Similarity 100.0%; Pred. No. 3.8e-167;  
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAVRALKLLTLLAVVAAASQAEVSEAGWGMVTPDLLFAEGTAAAYARGDPGVLSMER 60  
DB 1 MAVRALKLLTLLAVVAAASQAEVSEAGWGMVTPDLLFAEGTAAAYARGDPGVLSMER 60  
QY 61 ALRSRAALRALRLRCRTOCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
DB 61 ALRSRAALRALRLRCRTOCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
QY 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHFTFVGNPEHMEMQ 180  
DB 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHFTFVGNPEHMEMQ 180  
QY 181 QNLDDYOTMSGVKEADFKDLETQPHMQEFLRGVLYSEBEOQEAAPHLEAALQYFVAYE 240  
DB 181 QNLDDYOTMSGVKEADFKDLETQPHMQEFLRGVLYSEBEOQEAAPHLEAALQYFVAYE 240

QY 241 ECRALCEGPDYDYGYNLYEYNADLFQAITDHYIOVLNCKQNCVTELASHPSREKPFDFL 300  
DB 241 ECRALCEGPDYDYGYNLYEYNADLFQAITDHYIOVLNCKQNCVTELASHPSREKPFDFL 300  
QY 301 PSHYNYLOFAYYNGNTOAGECAKTYLLFPNDEVNMNOLAYVAALGEEHTRISIGPRE 360  
DB 301 PSHYNYLOFAYYNGNTOAGECAKTYLLFPNDEVNMNOLAYVAALGEEHTRISIGPRE 360  
RESULT 8  
US-09-833-245-1251  
; Sequence 1251, Application US/09833245  
; Publication No. US20040010134A1  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: Albumin Fusion Proteins  
; FILE REFERENCE: PFS46PCT  
; CURRENT APPLICATION NUMBER: US/09/833,245  
; CURRENT FILING DATE: 2001-04-12  
; PRIOR APPLICATION NUMBER: 60/229,358  
; PRIOR FILING DATE: 2000-04-12  
; PRIOR APPLICATION NUMBER: 60/256,931  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/199,384  
; PRIOR FILING DATE: 2000-04-25  
; NUMBER OF SEQ ID NOS: 2267  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1251  
; LENGTH: 359  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-833-245-1251

Query Match 46.5%; Score 1800; DB 11; Length 359;  
Best Local Similarity 99.4%; Pred. No. 6.4e-157;  
Matches 340; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 MAVRALKLLTLLAVVAAASQAEVSEAGWGMVTPDLLFAEGTAAAYARGDPGVLSMER 60  
DB 1 MAVRALKLLTLLAVVAAASQAEVSEAGWGMVTPDLLFAEGTAAAYARGDPGVLSMER 60  
QY 61 ALRSRAALRALRLRCRTOCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
DB 61 ALRSRAALRALRLRCRTOCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
QY 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHFTFVGNPEHMEMQ 180  
DB 121 RRCGLGPPAAHSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHFTFVGNPEHMEMQ 180  
QY 181 QNLDDYOTMSGVKEADFKDLETQPHMQEFLRGVLYSEBEOQEAAPHLEAALQYFVAYE 240  
DB 181 QNLDDYOTMSGVKEADFKDLETQPHMQEFLRGVLYSEBEOQEAAPHLEAALQYFVAYE 240  
QY 241 ECRALCEGPDYDYGYNLYEYNADLFQAITDHYIOVLNCKQNCVTELASHPSREKPFDFL 300  
DB 241 ECRALCEGPDYDYGYNLYEYNADLFQAITDHYIOVLNCKQNCVTELASHPSREKPFDFL 300  
QY 301 PSHYNYLOFAYYNGNTOAGECAKTYLLFPNDEVNMNOLA 342  
DB 301 PSHYNYLOFAYYNGNTOAGECAKTYLLFPNDEVNMNOLA 342

RESULT 9  
US-10-094-749-2560  
; Sequence 2560, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROFUKI  
; APPLICANT: ISHII, SHIZUKO



```

; SOFTWARE: PatentIn 2.1
; SEQ ID NO 24
; LENGTH: 736
; TYPE: PRN
; ORGANISM: Homo sapiens
US-10-094-886-24

Query Match
Best Local Similarity 42.1%; Score 1473; DB 15; Length 736;
Matches 321; Conservative 116; Mismatches 266; Indels 60; Gaps 12;

QY 6 LKLLTLLAVVAASQAQVES-----EAGGMVMTDLLFAEGTAAYARGDWPVVLSME 59
Db 2 LRLRLPLLLLLPPPPSGPPGLTQLSPGAPDPLLYADGLRAYAGANAPAVALLR 61
QY 60 RALRSRAALRALRLRCKTQCAAD-----FPWEL-----DPWSPSPAQAGALRDLSPFG 111
Db 62 EALRSQAALGRVLDGSCAADFGAALPAVLILGAPEDSGPGTQGSW-----ERQLLR 116
QY 112 GLLRRAACLRRC---LGPAAHSLSEMELE---FRKRSYNYLVQAYFKINKLEKAVAA 165
Db 117 AALRRADCLTQCAARRLPGGARLRVGSALRAFRRREPYNLYQAYQLKLLDLAANA 176
QY 166 AHTFFVGNPHMEMQNQLDYQTMGSKVEADFKDLETQPMQEFRLGVRLYSSEQPOEAV 225
Db 177 AHTFFVANPMHLQWRDMAKYRMSGVRPOSFRDLETPPHWAAYDTGLELLGRQEAAGLAL 236
QY 226 PHLEALQOEYFVAYEECRALCEGPDYDGYNYLEYNA----DLFOALTHYIOVLNCKON 281
Db 237 PRLEALQGSIAQWESCRACDEGPEEQGABEEDGAASOGGYEAYAGHWIOLVQCROR 296
QY 282 CVTELASHPSREKPFDFLPSHNYLQFAYINIGNTYQAGECAKTYLLFFPNDEVNQNQL 341
Db 297 CVGEAATPRGSPVPDFLNLRLRHEAQAQGNLSQAIENVLSVLLFFVEDEAAKRAL 356
QY 342 AYYAAMLGEHTSIGRESAKYRQBSLLEKELLFPAYDVGIFPVDPDSWTPPEVIEK 401
Db 357 NQYQAQGEPRP-GLGPREDIQFIRLSLGEKRLYYAMEHLGTSFKDPDPWTPPALIPE 415
QY 402 RLOEKOKSERETAVRISQETGNLMKETETLVEEKTESLDVSLRTREGGPLYEGISLTM 461
Db 416 ALREKLAREDEK-----RWDHEPVKPKPLTYWKVLLLEGVTLIQ 456
QY 462 NSKLLNGYQYVMDGVISDHECQELQRLTNVAATSG--DQYQGTSPHTPNEKPYGVTVF 519
Db 457 DSRQLNGSERAVLDGLTTPAECGVLLQLAKDAAGAGARSYGRRRSPHTPHERFEGTLVL 516
QY 520 KALKLQEGKVPLOSALHYNYNTEKVRIMESYERLDTPLYFSYSHLVCRATAEEVQAR 579
Db 517 KAAQLARAGTVGSGAKLLLEVSRVRLTQAYFSPERPLHLSTHLVCKRSALTEGEQOR 576
QY 580 KDSHPVHVDNCILNAETLVCVKEPPAYTPRDYSAILYLNGDFDGGNFYFTELDKATVTA 639
Db 577 MDLSHPVHADNCVLDPTGSCWREPPAYTYRDISGLLYLNDDFQGGDLPTTEPNALTVA 636
QY 640 EVQPOCGRANVFSGTENPHGVKAVTRGQRCATALWFLDPRHSERDRVQADLVKMLFS 699
Db 637 RVPRCGRVAFSGVENPHGVNAVTRGRCALAWHTWAPEREQEWIEAKELLQ---E 693
QY 700 PEEMDLQEQPLDAQQGPFPAPQ-----ESLSGSESCKPKDEL 736
Db 694 SQEEEBEEMPSKDFSPSPSRHRQORVQDKTRAPRVREEL 736

RESULT 11
US-10-012-697-1493
; Sequence 1493, Application US/10012697
; Publication No. US20030215803A1
; GENERAL INFORMATION:
; APPLICANT: Escobedo, Jaime
; APPLICANT: Garcia, Pablo Dominguez
; APPLICANT: Kassam, Altaf
; APPLICANT: Lamson, George
```

```

; APPLICANT: Scott, Beth
; APPLICANT: Drmanac, Radoje
; APPLICANT: Cirkvenjakov, Radomir
; APPLICANT: Dickson, Mark
; APPLICANT: Drmanac, Snezana
; APPLICANT: Labat, Ivan
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Kita, David
; APPLICANT: Garcia, Veronica
; APPLICANT: Jones, Lee William
; APPLICANT: Stache-Crain, Birgit
; TITLE OF INVENTION: HUMAN GENES AND GENE EXPRESSION PRODUCTS
; TITLE OF INVENTION: ISOLATED FROM HUMAN PROSTATE
; FILE REFERENCE: 2300-16252
; CURRENT APPLICATION NUMBER: US/10/012,697
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: 60/254,648
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: 60/275,668
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 1568
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1493
; LENGTH: 282
; TYPE: PRN
; ORGANISM: Homo sapiens
US-10-012-697-1493

Query Match
Best Local Similarity 23.4%; Score 906.5; DB 15; Length 282;
Matches 172; Conservative 48; Mismatches 62; Indels 23; Gaps 4;

QY 432 VBEKTESLDVSRITREGGPLYEGISLTMKSLNGYQYVMDGVISDHECQELQRLTN 491
Db 1 MEKRLSPKID--RDLREGGPLYENITFVNSEQLNGTQRLVLDNVLSEEQRELHVSAS 58
QY 492 VAATSGDGYRGTSHTPNEKPYGVTVFKALKLQEGKVPLOSALHYNYNTEKVRIMES 551
Db 59 GIMLVGDGYRGTSHTPNEKPYGVTVFKALKLQEGKVPLOSALHYNYNTEKVRIMES 118
QY 552 YERLDTPLYFSYSHLVCRATAEEVQARERKDDSHPVHVDNCILNAETLVCVKEPPAYTPRD 611
Db 119 YFNLNSTLYFSYTHMVCTALSGQDRNDLSHPHADNCLLDPEANECWKEPPAYTPRD 178
QY 612 XSAILYLNGDFDGGNFYFTELDKATVTAEVQPCGRANVFSGTENPHGVKAVTRGQRC 671
Db 179 YSAILYMNDDEFGGFIETEMDAKTVTASIKPKGRMISFSSGGENPHGVKAVTKGRCA 238
QY 672 IALWTLDRHSERDRVQADLVKMLFSPBEMDLQEQPLDAQQGPFPAPQESLSGSES 731
Db 239 VALWTLDRHSERDRVQADLVKMLFSPBEMDLQEQPLDAQQGPFPAPQESLSGSES 731
QY 732 PKDEL 736
Db 278 PKDEL 282

RESULT 12
US-09-764-868-736
; Sequence 736, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 736
; LENGTH: 267
; TYPE: PRN
```

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (6)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-736

Query Match          21.6%; Score 836.5; DB 9; Length 267;
Best Local Similarity 54.9%; Pred. No. 3.1e-68;
Matches 158; Conservative 45; Mismatches 64; Indels 21; Gaps 3

QY 449 GGPLYEGISLTMSKLLNGYQVRVMDGVYISDHCQELQRLTNVAATSGDGRGQTSPT 508
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 GTPCKXTGVRNSRVDPRVGRGTORVLLDNLVLSSEQRELHSAVGIMLVGDGVRGKTSPT 60

QY 509 PNEKPYGVTVFKALGLQEGKVPLQSAHLVYNVTEKVRIMESYFRIDTPLFYFSVSLVC 568
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 PNEKPEGATVLKALSGYEGRVPLKSAFLVDISEKARRIVESYFMLNSTLYFSYTHMVC 120

QY 569 RTAIEVQAEKRDSDHPVHVHVCILNAETLVCKVEPPAYTFRDYSALLYLNGDFDGGNFY 628
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 RTALSGQDQRNDLSHPVHAQNCULDPEANECKVEPPAYTFRDYSALLYMNDDFEGEPI 180

QY 629 FTFLDAKTVTAEVQPCGRAGVFSGGTNPENHGKAVTRGQRCACIALMFTLDPHSHSRDV 688
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 FTIEMDAKTVTASIKPKCGRMVLSFGSGGNPHGVKAVTKGKCAVALWFTLDPILYRELERI 240

QY 689 QADDLVKMLFSPSEMDLSQEQFLDAQQGGPPPAQESLGSSESXPKDEL 736
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 241 QADEVIAI-----LDOE-----OOCKHE-----LINPKDEL 267

```

```

RESULT 13
US-09-764-875-1106
; Sequence 1106, Application US/09764875
; Publication No. US20040018969A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ202
; CURRENT APPLICATION NUMBER: US/09/764,875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1106
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (6)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-875-1106

```

	Query Match	21.6%;	Score 836.5;	DB 11;	Length 267;
	Best Local Similarity	54.1%;	Pred. No. 3.1e-69;		
	Matches 158;	Conservative 45;	Mismatches 64;	Indels 21;	Gaps 3
QY	449	GGFLLYEGISLTMNSKLNGYQRVVMDSVSDHCEQLRLTNVAATSGDGYRGQTSPT	508		
Db	1	GTECKXTGVRSNRVDPVRVGRGTQTVLLDNVLSEEQRELHSHVSGIIMLVGDGYRGKTSPT	60		
QY	509	PNEKEVGVYTVFKALKIGOEKGVPLQSAHLYNNTYKVRITMESYFRLDPLTFYSYSLHVC	568		
Db	61	PNEKEGATVLKALSKSGYEGVRPKSARLFDISEKARRIVESYFWMLNTLFSYTHWC	120		
QY	569	RTAIEVQAEARKDQSHPVHVDNCILNATETLVCKVPPPAYTFRDYSAILYLINGDFGGNPFY	628		
Db	121	RTALSGQDQRNDLSHPITHADNCLLDPPEANCKEKPAYTFRDYSALLYMNDPFGGGFI	180		
QY	629	FTELDAKTVTAEVQPCQGRGVFSFGSTENPHGVKATRGQRCALALWFITLDPHESRDV	688		

```

Db      181 FTAMDAKVTIASIKPKGRMISFSSGGENPHGCVKAVTKGRCAVALWFTLDPLVRELERI 240
Qy      689 QADDLVKMLFSPPEMDLSQEQLDAQQGPPEPAQESLSGSESKPKDEL 736
      |||||: ||||| ||||| |||||
Db      241 QADEVIAI-----LDQE-----QQGKHE-----LNINPKDEL 267

RESULT 14
US-09-833-245-1250
; Sequence 1250, Application US/098333245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229, 358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256, 931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199, 384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1250
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-245-1250

```

Query Match	17.0%;	Score 657;	DB 11;	Length 173;
Best Local Similarity	99.2%;	Pred. No. 5.6e-52;		
Matches 128;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0
Qy	1	MAVRALKILTTLLAVAAAQAQEVESAGGWTTPDLLFAEGTAAAYARDGFWVLSW	60	
Db	1	MAVRALKILTTLLAVAAAQAQEVESAGGWTTPDLLFAEGTAAAYARDGFWVLSW	60	
Qy	61	ALRSAALRALRLRCRTCCAADPFWELDPDWSPPAQASGAGALRDLSPFGGLLRPAACL	120	
Db	61	ALRSAALRALRLRCRTCCAADPFWELDPDWSPPAQASGAGALRDLSPFGGLLRPAACL	120	
Qy	121	RRCLGPPAA	129	
Db	121	RRCLGPPAA	129	

```

RESULT 15
US-09-764-875-815
; Sequence 815, Application US/09764875
; Publication No. US20040018969A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ202
; CURRENT APPLICATION NUMBER: US/09/764,875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 815
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-875-815

```

Query Match 15.8%; Score 612; DB 11; Length 219;  
Best Local Similarity 61.2%; Pred. No. 1.2e-47;  
Matches 109; Conservative 31; Mismatches 38; Indels 0; Gaps 0

Qy 466 LNYGRVYMDGVI SDHCEQLRLTNVAATSGGVGQFSPHTPEKFGVTVFKALKLG 525



Db 4 VRTQVLLDNVLSEBQCRELHVSAGIMLVGDGYRGKTSPTPNKFEQATVLKALKSG 63  
QY 526 QEGKVPLQSAHLYYNVTEKVRIMESYFRLDTPLYFSYSHLVCKRTAIEEVQAEKDDSH 585  
Db 64 YEGRVPLKSAFLFYDISEKARRIVESYFMLNSTLYFSYTHMVCTALSGQDQRRNDL 123  
QY 586 VHVDCILNAETLVCKEPPAYTFRDYSAILLYLNGDFDGNFYFTELDARTVTAEVOP 643  
Db 124 IHADNCILLPEANEKWEKPPAYTFRDYSAILLYMDDFEGEFIFTEMADARTVTAIXP 181

Search completed: July 18, 2004, 09:56:56  
Job time : 687 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 09:30:43 ; Search time 57 Seconds  
(without alignments)  
1242.052 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 3870  
Sequence: 1 MAVRAKLKLTLLAVVAAS.....PPEPAQESLSGSESKPKDEL 736

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	500.5	12.9	431	2 A56822	synaptonemal compl
2	157.5	4.1	603	2 B86608	excinuclease ABC,
3	157.5	4.1	603	2 F72017	excinuclease ABC,
4	144	3.7	383	2 A96704	hypothetical prote
5	141	3.6	342	2 E70463	conserved hypothet
6	125.5	3.2	1033	2 T42701	hypothetical prote
7	124	3.2	1129	2 T43674	EGL-27 protein - C
8	123.5	3.2	640	2 F75114	hypothetical prote
9	122	3.2	1529	2 T02730	RNA-directed DNA p
10	120.5	3.1	2231	1 A46147	spectrin beta chai
11	118.5	3.1	1040	2 AH1926	hypothetical prote
12	118	3.0	1082	2 H81982	hypothetical prote
13	118	3.0	1597	2 S68420	citron - mouse
14	116.5	3.0	746	2 T19409	hypothetical prote
15	116.5	3.0	754	2 B88252	protein F44G4.1 [i
16	116	3.0	864	2 B90395	purine NTPase [imp
17	116	3.0	1166	2 A39432	ATP-dependent deox
18	115.5	3.0	1992	1 S02771	myosin heavy chain
19	115	3.0	866	2 I79267	trab protein - Esc
20	114.5	3.0	613	2 A35296	secretogranin II p
21	114	2.9	1922	2 T00637	hypothetical prote
22	113	2.9	1127	2 G71274	hypothetical prote
23	113	2.9	1805	1 A64224	hypothetical prote
24	112.5	2.9	726	2 F70355	penicillin binding
25	112.5	2.9	1287	2 T42658	hypothetical prote
26	111.5	2.9	722	2 T21854	hypothetical prote
27	111	2.9	2431	1 MNWVSF	nonstructural poly
28	110.5	2.9	1116	2 T16112	hypothetical prote
29	110	2.8	859	2 H70327	DNA mismatch repai

30	110	2.8	1788	2 T29043	hypothetical prote
31	109.5	2.8	543	2 A38093	transformation-sen
32	109.5	2.8	690	2 AG2419	hypothetical prote
33	109	2.8	1331	2 S75000	protoporphyrin IX
34	109	2.8	2157	2 A13009	peptide synthetase
35	109	2.8	2566	2 E98274	hypothetical prote
36	108.5	2.8	887	2 S70642	ubiquitin ligase N
37	108.5	2.8	1948	2 S00485	gene 11-1 protein
38	108	2.8	1179	2 G95144	conserved hypothet
39	108	2.8	2663	1 S28261	centromere protein
40	107.5	2.8	624	2 A55220	penicillin-binding
41	107.5	2.8	714	2 H64543	fumarate reductase
42	107.5	2.8	729	2 T50989	hypothetical prote
43	107.5	2.8	1965	2 T33216	hypothetical prote
44	107	2.8	829	2 F83905	hypothetical prote
45	107	2.8	1439	2 T47766	hypothetical prote

ALIGNMENTS

RESULT 1

A56822  
synaptonemal complex protein SC56 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 05-Jan-1996 #sequence\_revision 05-Jan-1996 #text\_change 04-Mar-2000  
C:Accession: A56822; S20742  
R:Chen, Q.; Pearlman, R.E.; Moens, P.B.  
Biochem. Cell Biol. 70, 1030-1038, 1992  
A:Title: Isolation and characterization of a cDNA encoding a synaptonemal complex protein  
A:Reference number: A56822; MUID:93213429; PMID:1363622  
A:Accession: A56822  
A>Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-431 <CHE>  
A:Cross-references: EMBL:X65454; NID:G57191; PIDN:CAA45449.1; PID:G57192  
A:Experimental source: testis  
A>Note: sequence extracted from NCBI backbone (NCBIP:129004)  
C:Superfamily: rat synaptonemal complex protein SC56

Query Match 12.9%; Score 500.5; DB 2; Length 431;  
Best Local Similarity 35.9%; Pred. No. 9.3e-27;  
Matches 124; Conservative 47; Mismatches 153; Indels 21; Gaps 8;

QY	46	YARGWPGVLSMERALRSRAALRLRCRTQCAADPPWELDPWSPSP----	AQASGAG	102
DB	14	VEGESWRNSARYLEALRLRLDSEAFCHANGSG--PATSQRPAPGPDGNEGDED	71	
QY	103	ALRDLSTFGGLLRRAACLRRCLGPPAAHSL---SEEMLEFRKSPNYLQVAYFKINKL	159	
DB	72	WARELRLFGHVLERAACLRRCRKTLPAPQVPYPSRQLLRDPQNLRYQLHYAHFKANRL	131	
QY	160	EKAVAAAHTEFVGNPHEMEMQNLDYQTMGSKVEADFKDLETOPHMQEERLGVRLYSEE	219	
DB	132	EKAVAAATFLQRPKHELTAKYLNYYHGMLDIGESITDLEAQFYEAFLQAVKLYNSG	191	
QY	220	QFQEAAPHLEAALQRYFYVAEBCRALCEGPDYDYGNYLYNADLFOAITDHYIQVLNCK	279	
DB	192	DFRSSTEHRALADYMTVFARCLAGCEGAHEQVDFK-----DFYPAIDLPAESLQCK	245	
QY	280	QNCVTELASHSPREKPPED-FLPSHYNYLQPAYNIGNYTOAGCAKTYLLFFPNDEVMN	338	
DB	246	VDCEANLT--ENVGGFFVDFKVFATYHYLQFAYKLVNDVHQAAASAASMYLFDPKDSVMQ	303	
QY	339	QNLAYV---AAMLGEHTRSGIPRESAKYQRSLLEKELFFAY	380	
DB	304	QNLVYVYRHRARWGLEE-EDQFPREEAVLYNQISELLELDFTH	347	

RESULT 2

B86608  
excinuclease ABC, subunit C [imported] - Chlamydomonas reinhardtii (strain J138)  
C:Species: Chlamydomonas reinhardtii (strain J138)

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C/Accession: B86608  
R:Shirai, M.; Hirakawa, H.; Kinoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ise,  
Nucleic Acids Res. 28, 2311-2314, 2000  
A/Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
A/Reference number: A86491; MUID:20330349; PMID:10871362  
A/Accession: B86608  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-603 <STO>  
A/Cross-references: GB:BA000008; NID:g8979314; PIDN:BA99148.1; GSPDB:GN00142  
A/Experimental source: strain J138  
C/Genetics:  
A/Gene: uvrC  
C/Superfamily: excinuclease ABC chain C

Query Match 4.1%; Score 157.5; DB 2; Length 603;  
Best Local Similarity 20.9%; Pred. No. 0.0068;  
Matches 131; Conservative 81; Mismatches 223; Indels 193; Gaps 27;

QY 51 WPGVLSMERALRGRALRALRL-----RCRTQCAADFPWELDPWSPSPAQASG 100  
DB 111 WPKV-----EAIRTKAITSSORQLIFGYPVSAEACHTLL-----EVSQWFP----- 152  
QY 101 AGALRDLSPFGGLLRRAC-----LRRCGLPPAAHSLSEMELEFRKSPYNYLQVAYFKI 156  
DB 153 ---LRTCSDFRALRKRPCILYDMKRLAPCVGYCTPEEYQGTLDK-----AIIFLK 201  
QY 157 NKLEKAVAAHAFTFVGNPEHMEMQNDLYCYTMSGVKEADFKDLETOPHMOEF----- 209  
DB 202 GKIEEVVKDLEKVIQKASDNLEFQAANYRTLSLIKQAMAKQOVKHFQNDALGLYR 261  
QY 210 -----RLGVRLYS--EQOQEAHPHLEAALQYFVA--YEECRALCEGYP 250  
DB 262 HKQRTILTLLTVRSGLKLGARHFFENAEQDQLSSFILQYVVSQYIPKPEILTPPL 321  
QY 251 DYDGVNLYENAD-----LFQAITDHYIOVL-----NCKQNCVTTELASHPSREKPFDF-- 299  
DB 322 EFPTLSYV--LNAESPRLRSPKTYGKELLDLAYRNKAYAAATL---PSSTLPYQDFQN 377  
QY 300 --LPSHYNYLOFAYYNIQNTQAGECAKTYLLFPNDENVNQNLAAYAAMLGEBHTRSIG 357  
DB 378 ILRMSQPYRIEYCDNA--HMQGAHATGVIVFENNGFDPKQ-----YRTFSI-- 423  
QY 358 PRESAGEYRQSLLEKELLFPAYDVGIPFVDPDSWTPEEVIKRLQEKOKSERETAVRI 417  
DB 424 --DSEKQNDLALLEEVL-----RRFSLTLPALPMIV-----VDGKTHYNTKKI 469  
QY 418 SOEIGNLMKEITELVEEKTES--LDVSRLLTREGGPLLYEGISLTMSKLLNGYQVRVMD 475  
DB 470 IQTNLNLTGIVQVTTAKESNHSRLNKEKIFCEFP---EGFSLPPTSNLLQFFQ----- 521  
QY 476 GVISDHECQELQRLTNVAATSGDGYRGQTSBHTPNEKFGYGVTVFKALKLGOEGKVPLOSA 535  
DB 522 -ILRD-----EAHRFAISKH---RKRGRKALFEQEKIPGIGEV----- 555  
QY 536 HLYYNVTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPVHVDNCILNA 595  
DB 556 -----KRLILQK-----FKSWKQVMLSSQEELEA----- 580  
QY 596 ETLVCVKEPPAYTPRDSAILYNGDFD 623  
DB 581 -----IPGLTKKDIAVLLARQKDFN 600

RESULT 3  
F72017  
excinuclease ABC, chain C CP0921.[imported] - Chlamydothila pneumoniae (strains CW1029 a  
C/Species: Chlamydothila pneumoniae, Chlamydia pneumoniae  
C/Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
C/Accession: F72017; E81522  
R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999

A/Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A/Reference number: A72000; MUID:99206606; PMID:10192388  
A/Accession: F72017  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-603 <ARN>  
A/Cross-references: GB:AE001673; GB:AE001363; NID:G4377252; PIDN:AA19078.1; PID:G4377252;  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, I.  
C.; Dodson, R.; Gwinn, M.; Nelson, M.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A/Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A/Reference number: A81500; MUID:20150255; PMID:10684935  
A/Accession: E81522  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-603 <REA>  
A/Cross-references: GB:AE002251; GB:AE002161; NID:G7189833; PIDN:AAF38706.1; PID:G7189833;  
A/Experimental source: strain AR39, HL cells  
C/Genetics:  
A/Gene: uvrC; CP0921  
C/Superfamily: excinuclease ABC chain C

Query Match 4.1%; Score 157.5; DB 2; Length 603;  
Best Local Similarity 20.9%; Pred. No. 0.0068;  
Matches 131; Conservative 81; Mismatches 223; Indels 193; Gaps 27;

QY 51 WPGVLSMERALRGRALRALRL-----RCRTQCAADFPWELDPWSPSPAQASG 100  
DB 111 WPKV-----EAIRTKAITSSORQLIFGYPVSAEACHTLL-----EVSQWFP----- 152  
QY 101 AGALRDLSPFGGLLRRAC-----LRRCGLPPAAHSLSEMELEFRKSPYNYLQVAYFKI 156  
DB 153 ---LRTCSDFRALRKRPCILYDMKRLAPCVGYCTPEEYQGTLDK-----AIIFLK 201  
QY 157 NKLEKAVAAHAFTFVGNPEHMEMQNDLYCYTMSGVKEADFKDLETOPHMOEF----- 209  
DB 202 GKIEEVVKDLEKVIQKASDNLEFQAANYRTLSLIKQAMAKQOVKHFQNDALGLYR 261  
QY 210 -----RLGVRLYS--EQOQEAHPHLEAALQYFVA--YEECRALCEGYP 250  
DB 262 HKQRTILTLLTVRSGLKLGARHFFENAEQDQLSSFILQYVVSQYIPKPEILTPPL 321  
QY 251 DYDGVNLYENAD-----LFQAITDHYIOVL-----NCKQNCVTTELASHPSREKPFDF-- 299  
DB 322 EFPTLSYV--LNAESPRLRSPKTYGKELLDLAYRNKAYAAATL---PSSTLPYQDFQN 377  
QY 300 --LPSHYNYLOFAYYNIQNTQAGECAKTYLLFPNDENVNQNLAAYAAMLGEBHTRSIG 357  
DB 378 ILRMSQPYRIEYCDNA--HMQGAHATGVIVFENNGFDPKQ-----YRTFSI-- 423  
QY 358 PRESAGEYRQSLLEKELLFPAYDVGIPFVDPDSWTPEEVIKRLQEKOKSERETAVRI 417  
DB 424 --DSEKQNDLALLEEVL-----RRFSLTLPALPMIV-----VDGKTHYNTKKI 469  
QY 418 SOEIGNLMKEITELVEEKTES--LDVSRLLTREGGPLLYEGISLTMSKLLNGYQVRVMD 475  
DB 470 IQTNLNLTGIVQVTTAKESNHSRLNKEKIFCEFP---EGFSLPPTSNLLQFFQ----- 521  
QY 476 GVISDHECQELQRLTNVAATSGDGYRGQTSBHTPNEKFGYGVTVFKALKLGOEGKVPLOSA 535  
DB 522 -ILRD-----EAHRFAISKH---RKRGRKALFEQEKIPGIGEV----- 555  
QY 536 HLYYNVTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPVHVDNCILNA 595  
DB 556 -----KRLILQK-----FKSWKQVMLSSQEELEA----- 580  
QY 596 ETLVCVKEPPAYTPRDSAILYNGDFD 623  
DB 581 -----IPGLTKKDIAVLLARQKDFN 600

RESULT 4

```

QY      196 D---FIDLTPQMQR--LGVRLYSBQPQAVPHL-BAALQEPV-----VAYECSRAL 245
DB      79 BBSFKALSINFSYSEARKNLGILYYKGRYEALKYLOEAANDYEYKKHEAFYYLAKV 138
QY      246 CEGPYDGY-NYLE----YNADLFOAITDHVIQVLNCKQCNCVTELASHPSREKPPEDFL 300
DB      139 YEAKQDLKNYVRYLEKAVYNPNFVA-----QLSLAQAYEN-----LGKYEEAEKIYKSLL 190
QY      301 PSHYN-VIQQ-----AYNIGNYTQAGCAKTYLLFFPNDENVNQNLAYYAAMLGGEHTR 354
DB      191 LNFGNFPFLKYLAEBVVYKGDYERAREIK-----BLLYKEN:TNEQ--- 233
QY      355 SIGPRESAKEYRQSLL--EKELLFPAYDFVGIPFPDPDSWTPEEVIPKRLOEQKOKSERE 412
DB      234 ----RKXVELLTKVLLAQORKLI-----IPRVHK-----PIKKEKK-KKY 271
QY      413 TAVRISOEI:GNLMKEITLVEE-KTKESLDVSRLTREG 449
DB      272 YAVQLG-AFSTKERADKLVLQELKSKGLRDRLILPTDG 307

RESULT 6
T42701
hypothetical protein DKFZp434G156.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
C;Accession: T42701
R;Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, November 1999
A;Reference number: Z22234
A;Accession: T42701
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1033 <AAA>
A;Cross-references: EMBL:AL133027
A;Experimental source: adult testis; clone DKFZp434G156
C;Genetics:
A;Note: DKFZp434G156.1

Query Match          3.2%; Score 125.5; DB 2; Length 1033;
Best Local Similarity 19.6%; Pred. No. 2.5;
Matches 97; Conservative 76; Mismatches 213; Indels 109; Gaps 20;

QY      132 LSEMELEFRKSPNYLVQVYPKINKLEKAV---AAAHFTFFVGNPEH---MEMQCNLDYY 186
DB      4 LEHEKSELKEIE-QELHLAQAEIOSLRQAADSATEHSEDIASIQEDLCRMQNLEDMM 61
QY      187 QTWSGVKEADFKDLETQPMQBRP---LGVRLYSBQPQAVPHLEAALQEFVAYEEC 242
DB      62 ERIRGYEMIASIRAMEMKSSPESSGLSDSYSGLOBE-----LQELPERHYFLNEEY 116
QY      243 RALCEGFYDGYNYLEYNADLFQAITDHVIQVLNCKQCNCVTELASHPSREKPPED--- 298
DB      117 RALQESNSSLTG-QLADLESERTHGATERWLQ---SQTLMTSAESOTSSEMDFLEPDPE 171
QY      299 --FLPSHYNVLPAYYNI GNYYTAGSCAKTYLLFFPNDENVNQNLAYYAAMLGGEHTRS 356
DB      172 MQLLRQOLRADEQEMGMKNKQCELCEBEL-----QHRRQV 209
QY      357 GPRESAKEYRQSLLLEKELLFFAYDFVGIPFPDPDSWTPEEVIPKRL-----QEK 406
DB      210 SEEEQRLQRELKCAQNEVL-----RFQTSHSVTVQNEELKSRCLTCLTKKYDTSODE 260
QY      407 QKSERETAVRISOEIGNLMKEITLVEEKTESLDVSRLTRREGPLLYEGISLTMSKLL 466
DB      261 QNELKNMQLOLQELTRLQPKVMKSTLVENQSEKEL-LCRLOKL-HQHQNVTC-EKEKLL 316
QY      467 NGYORVMDGVIDSHECOEQLRTNNVAATSGDGYRGOTSPHTNEKPFYGVTVFKALGLQG 526
DB      317 ERQQQLQEE--LOCHREA-ELQHLRDIVASPKSENEXDTETHAQOLQ----- 358
QY      527 EGKVPLOSALHYNNTEKVRIMESYRDLTPIFYSFYSHLVCTATAEEVQAEKDDSHPV 586

```

359 Db -----EMKQLYQASKDELERQKHMYDQLEQL-----LLCQLELKLKA-----SHPI 401  
587 QY HVD--NCILNAETLV 599  
402 Db PEDGKCKANKCDTLL 416

RESULT 7  
T43674  
EGL-27 protein - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 21-Jul-2000  
C:Accession: T43674  
R:Herman, M.A.; Ch'ng, Q.; Hettentbach, S.M.; Ratliff, T.M.; Kenyon, C.; Herman, R.K.  
Development 126, 1055-1064, 1999  
A:Title: EGL-27 is similar to a metastasis-associated factor and controls cell polarity  
A:Reference number: 222620; MUID:99128194; PMID:9927605  
A:Accession: T43674  
A:Status: Preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1129 <HER>  
A:Cross-references: EMBL:AF096618; NID:G4689162; PIDN:AAD27790.1; PID:G4689163  
C:Genetics:  
A:Gene: egl-27  
A:Map position: 2

Query Match 3.2%; Score 124; DB 2; Length 1129;  
Best Local Similarity 19.2%; Pred. No. 3.6;  
Matches 152; Conservative 97; Mismatches 270; Indels 274; Gaps 39;

QY 130 HSLSEMELEFRKSPNYLQVAFKINKLEKAVAAHTEFVGNPEHMEMQNLDDYQTM 189  
Db 89 YLRDSVFVEVSQNEPVIAAICGFKYTKRDVVVVKLTTRYFRAD-----DIPETS 138  
QY 190 SGVKEADFQKLETPHMQEFLGVRLY-SEEQPOEAVPHLEA-ALQGYFVAYECCALCE 247  
Db 139 LNLKMQERALEINPHLCPSLARELNFSELSQITQPSCLRGKCIYEVYKDVPHARTVAD 198  
QY 248 GPYDYGNY-LEYNADLFOAITDHY-IQVKNCKQNCVTBLASHPSREKPEFDFLPSHYN 305  
Db 199 FSLNDTFFFLCHYNQDSTKLASTHVAIRVGTSTFOATLPMAECVGDSDSDRDLELYREN 258  
QY 306 YLOPAYNIGNYQAGSCAKTYLL-----FFNDEVNQNLAAYAA 346  
Db 259 SISGEE--DYIKLARCITYILSGNHLDSQKNARSLOVSLMLDEALIQHRSYKI 316  
QY 347 --MLGEBHTRSI-----GPRESAKERYQ--RSL-----EKELLFFA 379  
Db 317 DDALSELNANDIILTTDVTNMTQDDAKKFAKIGKQLGNKPSRIHRELLPHSREQLSVY 376  
QY 380 YDVFQIFVDPDSW--TFEEVPIKR-----LQEKOKSERET----- 413  
Db 377 Y-----LMKKTPEATKQAAARRVNPSTSIKRPTEKVKASRPSTEVLDPSAS 425  
QY 414 -----AVRISQEIGN 423  
Db 426 ESDVENNGPSGRACHCYGAESKDWHHANGLLLTDCRLHYKKYQGLQRIANRPSQVPAC 485  
QY 424 LMKEIETLVEE-----TKESLDVSRITREGGPLLYEGISLTWN---SKLL 466  
Db 486 LFRKRSNDEEESGVRTRAGKEORRRTPSSMSETPDRRSPSTVSGNAPNLTAETPKKL 545  
QY 467 NGY-----QRVMDGVTSDEHCEQLQ-----LTNVAATSGDGYRGQTSPTHTN 510  
Db 546 NGSVKRAPKPLENGVINNVKESNSEEPASTPPPTPLTN-CLTNGH-----PESSITN 600  
QY 511 EKFGVTVFKALK-----GQBGKPLQSAHLYVYNTEKVRIMESYFLDTPLYF 561  
Db 601 ----GETISKRMKVPESYDDDDDEEGKMTIDEG-----DDMPV 637  
QY 562 SYHLVCRTAIEEVQAE-----RKDSHPVHVD-----NCILNAETLVCKEPPAYT---F 609

638 Db -LNGFKKEESVEEIKLELNGTIKKENG--VETDPTTLTSCSMEARNEVC--ETPAVSVSEI 692  
QY 610 RDSYAILYLNGDFDGGNFYFTELDAAKTVTAEVQP-----QCGRVGFSGSTENPHGVK 662  
Db 693 RDET-----NGE-----TNSDLKD-DENVEDPSPEDTFELGNSVEFE--TKNAMFVR 736  
QY 663 AVTR--GORCAI-----ALWFTLDPHSEDRVQADDLVKMLFSPPEMDLSQEQ- 709  
Db 737 STVRSCGPRCARTDLIFKIKVGVWEKSIKEKERKVHLQN--QRIQDSKVAIQNQI 794  
QY 710 PLDAQOQGPPEPAQ 722  
Db 795 KKEQOQSQTPTQQ 807

RESULT 8  
F75114  
Hypothetical protein PAB0498 - Pyrococcus abyssi (strain Orsay)  
C:Species: Pyrococcus abyssi  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
C:Accession: F75114  
R:Anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome struc  
A:Reference number: A75001  
A:Accession: F75114  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-640 <KAW>  
A:Cross-references: GB:A248285; GB:AL096836; NID:G5458067; PIDN:CAB49631.1; PID:el51552;  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: PAB0498

Query Match 3.2%; Score 123.5; DB 2; Length 640;  
Best Local Similarity 20.3%; Pred. No. 1.7;  
Matches 97; Conservative 68; Mismatches 140; Indels 174; Gaps 24;

QY 80 AADFPWE-----LDPDWSPPAQAGAGALRDLSPFGGLLRRAACLRCLGP--- 126  
Db 129 AAIMGEIRKDMVTGMINPDGSGPV-----GGILEKASAAHQAGAKIFL 174  
QY 127 -PAASHLSEMELEFRKSPNY-----YLOVAYFKI 156  
Db 175 IPEGRIQVVTETSKQIGPITQITTTQRRVDVAKYAQERWGLKVIIRDIDYDAVYFTG 234  
QY 157 NKLE-----KAVAAHTFFVGNPEHMEMQNLDDYQTMGSGVKEADFKOLETQPHMQEFLG 212  
Db 235 HKLERPKPKGVIVSTSFLEKYARTDYDETNNYN-----DVKO----- 273  
QY 213 VRLYSEEQPOEAVPHLEAALQGYFVAYECCR-ALCEGPDYDGYNLYEYNADLFOAITDH 271  
Db 274 -KLVNSDVGYSTYTLKALDDAYNLKQARDAIDQCKY-YTAMS-LDFQARIMRHVDW 330  
QY 272 YIQV-----LNCKQNCVTBL-----ASHPSREKPEFDFLPSHY 304  
Db 331 YIDVYDGRDIEDVFFVEKNEINSVEKYSNLTIKGVTMLQAVAASETRIEAEKYLKK-- 388  
QY 305 NYLOPAYN-----IGNYTOAGCAKTYLLFFPNDENVNQNLAYYAAMLGEEHTR-SIG 357  
Db 389 --AQASYNQDYWAVGNAAYAYERAKTAEIWN-----AKLGEFPAKSII 431  
QY 358 PRESAKERYQSLLEKELL-FFAYDVFG-----IPFVDPDSWTPEEVIKELQEKOK- 408  
Db 432 SRDSIKKAAREQLDNAKLIVNVTSMFGQQNLNLDMLDLKG-----ERYVEDGY 482  
QY 409 -----SEETAVR---ISQETG-----NLMKEIETLVEEKTESLDVSRITREG--PLL 453  
Db 483 SAALFSAMEARIRGEVILDTIGIENSVLRKXLSMMKEEAKTAIG---LAQEGTLPFL 538

RESULT 9  
T02730



Db 1456 EPIVVKTAFLRFEKIKAPL-----LERQKALEKKKEAFQCRD---VEDEK 1500

QY 598 LVCVKPEPPAYTFRDYSGAILYNGDFDGGNFYFTFELDAKTVTAEVQPCGRAGVFSSGTEN 657  
      :

Db 1501 LWDIEKLVPANSPDYGNLSLF-----NVHVLKKNQSLATEID-----NHE 1540

QY 658 PHGVKAVTRGQRCAIALWFLDPRHSERDVQA--DDLV-----KMLFSPE 701  
      :

Db 1541 PR-INAIICNNGRKLI-----DSGHDAKAFELISDLTKQWQELXDATENRRKHLESE 1593

QY 702 EMDLSQEQPLDAQQCPPEPAQESL--SGSSSKPKDEL 736

Db 1594 KV--QOYFDAQEASWSEQUELYMVMVEDRGKDEI 1626  
      :

RESULT 11

AH1926

hypothetical protein alr0963 [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002

C;Accession: AH1926

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Kikuchi, S.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 2005-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AH1926

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1040 <XUR>

A;Cross-references: GB:BA000019; PIDN:BA072920.1; PID:gl7130309; GSPDB:GNC00179

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: alr0963

```

+++      :       |       +       |||     ++      :   :|||    ++++
599 YLVIELLSKADSNLENLNPEDISNTRFABIKULLDDETVIIQWYIFDCFRAFIISKQHQP 658
QY -----COELQRLETNVAATSGDGVRGOTSPHTPNKEFYGVTVFK-AUKLGOGKVPLQSA 535
Db         ::||::++|||||+++++-----QIYGDEKKWRVQLNEQLTQTLT 704
QY 536 HLYYNVTWKPRIMESYRFLDTPLYFSYSHLCVRTAIEEVCQAERKDSDHPVVHVNCLINA 595
Db 705 HL-----NQITSLIPESQQVKGLIVIPHRYLHLF---PLHAVPLANNNSSQPEYLDFRPFHG 756
QY 596 ETLVCVGKPAPTFFDYSAILYYLN-GPDFG-----GNFYFTELDAKTVTAEVQPO 644
Db 757 VSYAPSNQLLRFTQRRVRRLANLELPNSULFAIQNTNDLAFTDIETVEITAADEFQO 814

RESULT 12
H81982
hypothetical protein NMA0631 [imported] - Neisseria meningitidis (strain Z2491 serogroup
C)Species: Neisseria meningitidis
C>Date: 05-May-2000 #sequence_revission 05-May-2000 #text_change 02-feb-2001
CAccession: H81982
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A>Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A'Reference number: AB1775; PMID:20223556; PMID:10761919
A'Accession: H81982
A>Status: preliminary
A'Molecule type: DNA
A'Residues: 1-1082 <PAR>
A'Cross-references: GB:AUI62753; GB:AUI57959; NTID:g73739120; PIDN:CAB83920.1; PID:g73739361
A'Experimental source: serogroup A, strain Z2491
C'Genetics:
A'Gene: NMA0631
C'Superfamily: Neisseria meningitidis hypothetical protein NMA0631
```



QY 325 -----KTYLLFPNDVNM-----QNLAYAAMLGBHTSICP-----RE 360  
Db 449 YGDHYGKNTTEKIXILFPADBIKPVVLRALSQARKVINGVVRVYSGPARIHIETARE 508  
QY 361 SAKEYRORSLEKELLFFAYDVFDPDSTPEVTPKRLQEKQKSERETAVRISOE 420  
Db 509 VGSFKDKRKEIE-----KROEENRKDRKAAAFREY 540  
QY 421 IGNLMKEIETVBEKTESLDVRLTRBGGPLLYEGISLTMSKLLNGVQVVMGVISD 480  
Db 541 FPNFVG-----BPKSKDILKRLYEQQHGKCLYSKKEINLGRNLKGYVEI-----D 587  
QY 481 HECQELQRLTNVAATSGDVGRTGTSPTNPNEKFGYGVTFKALKLQEGK-VPLQSAHLYX 539  
Db 588 HALP-----FRTWDDSENN-----KVLVLSGNQNKGNQTPYEYF 623  
QY 540 NVTEKVRIMESYFRDIT 557  
Db 624 NGKDNSEWQEFKARVET 641

RESULT 13  
S68420  
citron - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 15-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 05-Nov-1999  
C:Accession: S68420  
R:Madaule, P.; Furuyashiki, T.; Reid, T.; Ishizaki, T.; Watanabe, G.; Morii, N.; Narumiya  
FEBS Lett. 377, 243-248, 1995  
A:Title: A novel partner for the GTP-bound forms of rho and rac.  
A:Reference number: S68420; PMID:96128239; PMID:8543060  
A:Accession: S68420  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-1597 <MAD>  
A:Cross-references: EMBL:U93904; NID:g1079733; PID:AACS2341.1; PID:g1079734  
C:Superfamily: protein kinase C zinc-binding repeat homology; pleckstrin repeat homology  
C:Keywords: alternative splicing  
F;931-979/Domain: protein kinase C zinc-binding repeat homology <K22>

Query Match 3.0%; Score 118; DB 2; Length 1597;  
Best Local Similarity 20.5%; Pred. No. 16;  
Matches 120; Conservative 88; Mismatches 190; Indels 186; Gaps 28;

QY 132 LSBEMLFEFRKRPYNYLOVAYFKINKLEKAVAAHTFFVGN---PEHM---EMQQLDYY 186  
Db 343 LSEKAMINAMDSKIRLEQRIVELSEANKLAANSLSFTORNKKAQEBMISELRQOKFY 402  
QY 187 QTMGKVEADFKDLET-----OPHMQEFL-----GVPLYSEEQPQ---AVPHL 228  
Db 403 ETQAGKLEAQRKLEQLEKISHQDSKSRLLLETRLEVSLEHEEQKLEKRLQTEL 462  
QY 229 EAAALQ---YFVAYECCALCEGPDYDGYNYLEYNADL---FQAITDHYIQLN---C 278  
Db 463 QLSIQEESQLTALQARAALQSLQKQAKTELBETTAZAEETIQAHTARDEIQKFDAL 522  
QY 279 KQNC--VTEL-----ASHPSRKPFDF-----LPSHNYLOFAY- 311  
Db 523 RNSCTWITDLEQLNQLTDEANLNQNFYLSKQLDEASGANDEIVQLRSEVDHLRREIT 582  
QY 312 ---YINQNYTQAGCACTYLLFFP-----NDEVMONLAYA--AMLGEHT----- 353  
Db 583 EREMQLTSQKQWALKTCTMLEEQVLDLEALNDELLEKEROEWARSVLGDEKQSFEC 642  
QY 354 -----RSIGPRE-----SAKEYR-----QRSLEKELLFFAYDV 382  
Db 643 RVRELQMLDTEKQSRARADQRTESQVVELAVKEHKABILAQALQALKEQKL----- 695  
QY 383 FGIPFVDPDSWTP-----EVI PKRLQEKQKSERETAVRISOEIGNLMKEIET- 430  
Db 696 -----KAESLSDKLNLEKKHAWLENMARNSLQOKJETERELKQRLLEQAKLQQQMDLQ 749  
QY 431 -----LVEEKTBSLDSVRLTR-EGGPLLYEGISLTMSKLLNGVQVVMGVISDHEQC 484

## RESULT 14

T19409

hypothetical protein F44G4.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C&gt;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T19409; T22212

R:Simms, M.

submitted to the EMBL Data Library, March 1996

A:Reference number: Z19120

A:Accession: T19409

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-746 &lt;WIL&gt;

A:Cross-references: EMBL:Z70034; PIDN:CAA93858.2; GSPDB:GN00020; CESP:F44G4.1

A:Experimental source: Clone C18E9

R:Simms, M.

submitted to the EMBL Data Library, June 1995

A:Reference number: Z19530

A:Accession: T22212

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-746 &lt;W12&gt;

A:Cross-references: EMBL:Z49910; PIDN:CAA90124.2; GSPDB:GN00020; CESP:F44G4.1

A:Experimental source: Clone F44G4

C:Genetics:

A:Gene: CESP:F44G4.1

A:Map position: 2

A:Introns: 26/1; 62/1; 123/2; 326/2; 343/3; 532/3; 630/3; 654/2; 690/2; 733/3

Query Match

Best Local Similarity 3.0%; Score 116.5; DB 2; Length 746;

Matches 121; Conservative 106; Mismatches 239; Indels 161; Gaps 32;

QY 134 EMELEFRKR--SPNYLQVAYFKINKLEKAVAAHTFFVGNPEHM---EMQQLDYYOTM 189

Db 147 EBYQVELNRILTTVQKQORSIFTFDEADKL-----PEQLLGAIKPFLDYVSTI 195

QY 190 SGV--KEADFKDLETQPHMQEFLGVRLYSEQQAQVPHLEAALQVYFVAYECCALCE 247

Db 196 SGVDFRRSIFILLSNKGGGEIARITKQYESGYPREQL-RLEAFERELM----- 243

QY 248 GPYDYGNYLEYNADLFOAITDHYIQVLNCKQNCVTELASHPSREK-----PFEDFLP 301

Db 244 -NFSYNEKGLQMSLISNHLIDHVPFLPLQREHVRSCVGYLRKRGDGLSVNVDFVE 302

QY 302 SHYNYLQFAYNIGNYTOAGCACTYLLFFPNDVNMONLAYAAMLGEEHTRSGPRES 361

Db 303 RVLSNQYFPSSKAFSSSG-CKRV-----DAKTDLEMAKIRELLS-----SWAPKKS 349

QY 362 AKEYRQSLLEKELLFFAYDVFDPDGSW-----TPEEVI PKRLQEKQKSERETAVR 416

Db 350 KSKKEESMID-----FVEEVTG--DVDEDFQEAEDMPDEVDSEDFPKKKVVKKE 403

QY 417 ISQETIGNLMKEIETIVEE-----MTKSLDSVRLTR-EGGPLLYEGISLTMSKLL 466

Db 404 IKQBELTDEKLELLEKYEAKATATKTD--DPKHLPKSGRGKALR--ALRKDKRAR 459

QY 467 NGYQVVMGVISDHECOE-----LQRLTNVAATSGDVGRTGTSPTNPNEKFGYGVTFKAL 522

Db 460 QGERAQIRDELGESAPQKEVPKTIESMREYDATMVNEEDDEVEHDEANDEF----- 510

QY 523 KLGQEGKVPLOSAHLYN--VTEKVRIMESYFRLDTPLY-----FSYSHLV 567  
Db 511 -----AP-----YFNRETSPKVMITMTPKAKITTFKCFEQLQKCIPINSEIFTRKNVL 557  
QY 568 CRTAIEVQAEKDDSHPHVDNCILNAETLVCVKE-PPAY-----TPRDYSAILYLNG 620  
Db 558 LKTIIEQAKEREFTDLLVVEDRKKPNGIIFCHLPEGTAYFKINSLTFTQDLKVCY--- 614  
QY 621 DFDGNGFYFTELDAKTVAEVQPCQGRAVG--FSSGTENPHGVKAV-----TR-QORCA- 671  
Db 615 -FD--NFFM-----YCLSKSLKLFYKFGESTSHFPEVILNNTNRLGHNIAR 657  
QY 672 -IALWFTLDPR-----HSERDRV 688  
Db 658 MLACLPHDPKFTGRRVVTFFHNRDYI 684

RESULT 15  
B88252  
protein F44G4.1 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
C:Accession: B88252  
R:anonymous, the C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog  
A:Reference number: A75000; PMID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.ele  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999, and  
A:Accession: B88252  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-754 <STO>  
A:Cross-references: GB:chr\_II; PIDN:CAA90124.1; PID:g3877079; GSPDB:GN00020; CESP:F44G4.  
A:Gene: F44G4.1  
A:Map position: 2

Query Match 3.0%; Score 116.5; DB 2; Length 754;  
Best Local Similarity 19.3%; Pred. No. 6.4;  
Matches 121; Conservative 106; Mismatches 239; Indels 161; Gaps 32;

QY 134 EEMELERKR--SPYNYLVAYKINKLEKAVAAHTFFVGNPEHM--EMQQNLDDYYQTM 189  
Db 155 EYQVELNRILTVQCKQSIFIFDEADKL-----PEQLGAIKPELDYISTY 203  
QY 190 SGV--KEADFKDLETPHMOEFLRGLVLYSEEQPEAVPHLEAALQBYFVAYECCALCE 247  
Db 204 SGVDFRRSIFILLNKGGEIARITKEQYBSGYPREQL-RLEAFERELM----- 251  
QY 248 GPYDYGNYVLEYNADLFQALITHYIQVLNCKQNCYTELASHPSREK-----PFEDFLP 301  
Db 252 -NFSYNEKGGLOMSELISNHLIDHFVFLPQREHVRSCVGAYLRKRGDLVSNVDFVE 310  
QY 302 SHYNYLOFAYNIGNYQAGECAKTYLLFFPNDEVWQNLAHYAAMLGBEHTRSIGIPRES 361  
Db 311 RVLSLQYFPSSKAFSSG-CKRV-----DAKTDLEMAKIRPLLS-----SMAPKKS 357  
QY 362 AKEYRQSLLEKELLFFAYDVFGIPFVDPDSW-----TPEEVIKRLQEKQKSERETAVR 416  
Db 358 KKKKKEESMID----FYEEVETG--DVDEDFGEQAEDEMDDEDEPPKPKKKKYVKKE 411  
QY 417 ISQIGNLMKEIETLVPE-----KTKESLDVSRLTR-EGGPLLYEGISLTMSKLL 466  
Db 412 IKQELLETDEKLQELLEKEYSKATATKTD--DFKHLPKSQGKALKR--ALRKDKRAR 467  
QY 467 NGQVRVMDGVIDSHECQE-----LQRLTNVAATSGDGYRGQTSPHTPNKFKYGVTVFKAL 522  
Db 468 QGERAQIRDEBLGSAPQKQKPKIESNREYDATMVNEEDEDEVEHDEANDEF----- 518  
QY 523 KLGQEGKVPLOSAHLYN--VTEKVRIMESYFRLDTPLY-----FSYSHLV 567  
Db 519 -----AP-----YFNRETSPKVMITMTPKAKITTFKCFEQLQKCIPINSEIFTRKNVL 565

QY 568 CRTAIEVQAEKDDSHPHVDNCILNAETLVCVKE-PPAY-----TPRDYSAILYLNG 620  
Db 566 LKTIIEQAKEREFTDLLVVEDRKKPNGIIFCHLPEGTAYFKINSLTFTQDLKVCY--- 622  
QY 621 DFDGNGFYFTELDAKTVAEVQPCQGRAVG--FSSGTENPHGVKAV-----TR-QORCA- 671  
Db 623 -FD--NFFM-----YCLSKSLKLFYKFGESTSHFPEVILNNTNRLGHNIAR 665  
QY 672 -IALWFTLDPR-----HSERDRV 688  
Db 666 MLACLPHDPKFTGRRVVTFFHNRDYI 692

Search completed: July 18, 2004, 09:44:06  
Job time : 61 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 09:40:18 ; Search time 58 Seconds  
(without alignments)  
655.116 Million cell updates/sec

Title: US-10-045-815-4

Perfect score: 3870

Sequence: 1 MAVRAKLLTLLAVVAAS.....PPEPAQESLSGSSEKPKDEL 736

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/prodata/2/iaa/5A COMB.pap.\*
- 2: /cgn2\_6/prodata/2/iaa/5B COMB.pap.\*
- 3: /cgn2\_6/prodata/2/iaa/6A COMB.pap.\*
- 4: /cgn2\_6/prodata/2/iaa/6B COMB.pap.\*
- 5: /cgn2\_6/prodata/2/iaa/6C COMB.pap.\*
- 6: /cgn2\_6/prodata/2/iaa/6D COMB.pap.\*
- 7: /cgn2\_6/prodata/2/iaa/6E COMB.pap.\*
- 8: /cgn2\_6/prodata/2/iaa/6F COMB.pap.\*
- 9: /cgn2\_6/prodata/2/iaa/6G COMB.pap.\*
- 10: /cgn2\_6/prodata/2/iaa/6H COMB.pap.\*
- 11: /cgn2\_6/prodata/2/iaa/6I COMB.pap.\*
- 12: /cgn2\_6/prodata/2/iaa/6J COMB.pap.\*
- 13: /cgn2\_6/prodata/2/iaa/6K COMB.pap.\*
- 14: /cgn2\_6/prodata/2/iaa/6L COMB.pap.\*
- 15: /cgn2\_6/prodata/2/iaa/6M COMB.pap.\*
- 16: /cgn2\_6/prodata/2/iaa/6N COMB.pap.\*
- 17: /cgn2\_6/prodata/2/iaa/6O COMB.pap.\*
- 18: /cgn2\_6/prodata/2/iaa/6P COMB.pap.\*
- 19: /cgn2\_6/prodata/2/iaa/6Q COMB.pap.\*
- 20: /cgn2\_6/prodata/2/iaa/6R COMB.pap.\*
- 21: /cgn2\_6/prodata/2/iaa/6S COMB.pap.\*
- 22: /cgn2\_6/prodata/2/iaa/6T COMB.pap.\*
- 23: /cgn2\_6/prodata/2/iaa/6U COMB.pap.\*
- 24: /cgn2\_6/prodata/2/iaa/6V COMB.pap.\*
- 25: /cgn2\_6/prodata/2/iaa/6W COMB.pap.\*
- 26: /cgn2\_6/prodata/2/iaa/6X COMB.pap.\*
- 27: /cgn2\_6/prodata/2/iaa/6Y COMB.pap.\*
- 28: /cgn2\_6/prodata/2/iaa/6Z COMB.pap.\*
- 29: /cgn2\_6/prodata/2/iaa/6A COMB.pap.\*
- 30: /cgn2\_6/prodata/2/iaa/6B COMB.pap.\*
- 31: /cgn2\_6/prodata/2/iaa/6C COMB.pap.\*
- 32: /cgn2\_6/prodata/2/iaa/6D COMB.pap.\*
- 33: /cgn2\_6/prodata/2/iaa/6E COMB.pap.\*
- 34: /cgn2\_6/prodata/2/iaa/6F COMB.pap.\*
- 35: /cgn2\_6/prodata/2/iaa/6G COMB.pap.\*
- 36: /cgn2\_6/prodata/2/iaa/6H COMB.pap.\*
- 37: /cgn2\_6/prodata/2/iaa/6I COMB.pap.\*
- 38: /cgn2\_6/prodata/2/iaa/6J COMB.pap.\*
- 39: /cgn2\_6/prodata/2/iaa/6K COMB.pap.\*
- 40: /cgn2\_6/prodata/2/iaa/6L COMB.pap.\*
- 41: /cgn2\_6/prodata/2/iaa/6M COMB.pap.\*
- 42: /cgn2\_6/prodata/2/iaa/6N COMB.pap.\*
- 43: /cgn2\_6/prodata/2/iaa/6O COMB.pap.\*
- 44: /cgn2\_6/prodata/2/iaa/6P COMB.pap.\*
- 45: /cgn2\_6/prodata/2/iaa/6Q COMB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148.5	3.8	603	4	US-09-198-452A-1011
2	118	3.0	503	4	US-09-134-001C-3480
3	116	3.0	478	4	US-09-738-946-6
4	116	3.0	553	4	US-09-134-000C-4071
5	115	3.0	325	4	US-09-972-784-4
6	114	2.9	3878	4	US-09-914-259-11
7	111	2.9	2431	1	US-07-920-281C-2
8	111	2.9	2431	3	US-08-466-277-2
9	110.5	2.9	1257	1	US-08-049-783-2
10	110.5	2.9	1257	1	US-08-158-232-6
11	110.5	2.9	1257	1	US-08-304-626-6
12	110.5	2.9	1257	1	US-08-316-301A-6
13	110.5	2.9	1257	2	US-08-611-328-6
14	110.5	2.9	1257	3	US-08-073-891-6
15	110.5	2.9	1257	3	US-08-076-137-6
16	110.5	2.9	1257	5	PCT-US92-0362A-6
17	110.5	2.9	1257	5	PCT-US92-0362A-6
18	109	2.8	927	3	US-08-895-601-6
19	108.5	2.8	451	4	US-09-134-000C-3849
20	108	2.8	657	4	US-09-284-768A-7
21	108	2.8	1596	4	US-08-978-277A-4
22	108	2.8	2662	4	US-09-595-684B-31
23	106	2.7	608	4	US-09-284-768A-4
24	105	2.7	427	4	US-09-134-000C-4847
25	105	2.7	723	4	US-09-434-408-2
26	105	2.7	1122	4	US-09-252-991A-22843
27	105	2.7	3248	1	US-08-353-700-1

#### ALIGNMENTS

##### RESULT 1

US-09-198-452A-1011

; Sequence 1011. Application US/09198452A

; Patent No. 6559294

; GENERAL INFORMATION:

; APPLICANT: Grifffais, R.

; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention of infection and treatment of infection

; FILE REFERENCE: 9710-003-999

; CURRENT APPLICATION NUMBER: US/09/198,452A

; CURRENT FILING DATE: 1998-11-24

; NUMBER OF SEQ ID NOS: 6849

; SEQ ID NO 1011

; LENGTH: 603

; TYPE: PRT

; ORGANISM: Chlamydia pneumoniae

; US-09-198-452A-1011

Query Match 3.8%; Score 148.5; DB 4; Length 603;  
Best Local Similarity 20.8%; Pred. No. 1.3e-05;  
Matches 130; Conservative 79; Mismatches 229; Indels 187; Gaps 26;

28	105	2.7	3248	5	PCT-US95-16216-1	Sequence 1, Appli
29	104.5	2.7	662	4	US-09-134-000C-5682	Sequence 5682, Ap
30	104	2.7	1071	2	US-08-975-527-1	Sequence 1, Appli
31	104	2.7	1151	4	US-09-134-001C-3242	Sequence 3242, Ap
32	103.5	2.7	525	3	US-09-113-750A-35	Sequence 35, Appl
33	103.5	2.7	911	2	US-08-484-438-10	Sequence 10, Appl
34	103	2.7	1298	4	US-09-540-236-2334	Sequence 2334, Ap
35	102.5	2.6	368	4	US-09-328-352-4661	Sequence 4661, Ap
36	102.5	2.6	513	4	US-09-134-001C-4629	Sequence 4629, Ap
37	102.5	2.6	564	4	US-09-252-991A-20698	Sequence 20698, A
38	102	2.6	906	3	US-08-630-916A-48	Sequence 48, Appl
39	102	2.6	1485	4	US-09-543-681A-5367	Sequence 5367, Ap
40	101.5	2.6	599	3	US-08-556-419-22	Sequence 22, Appl
41	101.5	2.6	610	4	US-09-800-170-22	Sequence 22, Appl
42	101.5	2.6	732	4	US-09-307-143-4	Sequence 4, Appli
43	101	2.6	2482	1	US-08-328-254-6	Sequence 6, Appli
44	100.5	2.6	316	4	US-09-241-750-2	Sequence 2, Appli
45	100.5	2.6	1151	4	US-09-328-352-4744	Sequence 4744, Ap



; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; CURRENT FILING DATE: 2001-10-04
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 4071
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-4071

Query Match 3.0%; Score 116; DB 4; Length 553;
Best Local Similarity 18.6%; Pred. No. 0.02;
Matches 73; Conservative 67; Mismatches 119; Indels 134; Gaps 20;

QY 169 FVGNPE--HMEMQNLQYQM-----SGVKEADF-----KDELTQPHMOER 210
DB 208 FYTGKEDLYHKLKQATDYQVIDTGKAVASRIKEDVRFGYDLPKDAFQPTNAYK- 266

QY 211 LGVRLYSEBQPOBAPVHLEAALQEFVAYECCALCEGPDYDGYNLYEYNADLFQA--- 267
DB 267 LTVELSVKEM-----APFSWDSFALIQGETKAFEGSLL 299

QY 268 -----ITDHYIOVLNCKQNCVTELASHPSREKPEDFLPSHYNYLPAYYNYGNYTQA 320
DB 300 AQPATNEMENEFQV-KIENNGSLATADKKTGE-----TFSKLLTFEDT 342

QY 321 GECATYLLFFFNDE--VNNQNLAAYAAMLGBEHTRSIGPESAKERYORSLEKELIFF 378
DB 343 GDIGNEYIFKPTEDOGIITENVT--AEITNKENS-----PVKASQIKQTVMLP----- 390

QY 379 AYDVFGIPFVDPDSWTPEEVIKRLQEKOKSRETAVRISQIBGNLMK-EIETLVEEKTK 437
DB 391 -----VAAD-----ERLEEQKAVREFRERLAQSTTLRPFETIMV----- 427

QY 438 ESLDVSRLTREGGPLYEGISLTMSKLLNGYQVVM-DGVISD-HECCOELQRLTNVAAT 495
DB 428 -----TMIKESNQLFFE--TTINQIKDHLRLVLFPTGMVTHEADSIYEVVTRPQ 478

QY 496 SGDGYRGQTSPP-----HTPNEKFGYTVF 519
DB 479 VSDTWENPTNPQHQAFVNVHDQNK---GVTF 508

RESULT 5
US-09-972-784-4
; Sequence 4, Application US/09972784
; Patent No. 6566088
; GENERAL INFORMATION:
; APPLICANT: Bruik, Richard K.
; APPLICANT: McKnight, Steven L.
; TITLE OF INVENTION: Prolyl-4-Hydroxylases
; FILE REFERENCE: UTSD0871
; CURRENT APPLICATION NUMBER: US/09/972,784
; CURRENT FILING DATE: 2001-10-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 4
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-972-784-4

Query Match 3.0%; Score 115; DB 4; Length 325;
Best Local Similarity 22.3%; Pred. No. 0.01;
Matches 75; Conservative 44; Mismatches 121; Indels 96; Gaps 16;

QY 433 EETKESLD-----VSLRTREGGPLYEGISLTMSKLLNGYQVVM-DGVIS 479
DB 23 EQVFRELLDKRERYEDLCRNISDMNQVGLSVDDFLGMETGLKILNEVRSMYNAGAFQ 82

QY 480 DHECOELQRLTN-----VAATSGDGYRGQ-----TSPTHTPNEKF-----XGVTYVFKALKL 524
DB 83 DG-----QVVTNQTDPAPAVRGDKIRGDKIKWVGNEPGCSNVWYLTNQIDSVVYRVNTM 137

QY 525 QGEGKVPLOSALHYNYNTEKVRIMESYFRDLTPLYFSYSHLVCTALIEEVQAEKDDSH 584
DB 138 KDNGLG-----NYHIRERTRAMVAC-----PGSGTHIV----- 167

QY 585 FVHDNCILNAETLVCVKPEPAYTFRDYSAILYLNGDFD-----GNGFYFTELDARTYTAE 640
DB 168 -MHVDN-----PKDGRVITAIYILNINWDARESGLIRIRPTGTTV-AD 211

QY 641 VPOCGRAVGFSSTENPHGVKAVTRGQRCALIAFWLTPDPRHSERDRVQADDLVKMLFSP 700
DB 212 IEPFDFRLIFPWSDIRPHEVQPAHR-TRYAITVWY-FDAKERBEALIRA-----KLENSK 265

QY 701 EEMDLSEBQPLDAQGPPE-----PAQESLSGSESKP 732
DB 266 TNNLAAQAO-----AQAEFDSITTPPAAPASSASLIP 298

RESULT 6
US-09-914-259-11
; Sequence 11, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-914-259-11

Query Match 2.9%; Score 114; DB 4; Length 3878;
Best Local Similarity 19.0%; Pred. No. 0.95;
Matches 127; Conservative 95; Mismatches 258; Indels 190; Gaps 28;

QY 179 MQONLDYQTMGVKEADFQDLETPHMOERLGRVRLYSEBQPOEAVP-----HLEALQE 234
DB 2952 LEERKAYINTISL-----KDLITKMLQR---EAEVYDSSQSHESFSDWRGELLALQ 3003

QY 235 YFVAYECCALCEGPDYDGYNLYEYNADLFQAITDHYIOVLNCKQNCVTE-----LASH 289
DB 3004 VFL--EERSVLLAA-----FRTELTAQTDAVGLNCLCQRIQEGVEYQAAM 3050

QY 290 PSREKPFEDFLPSHYNYLPAYYNYGNYT--QAGECAKTYLLFPDENVNQNLAAYAA 347
DB 3051 ECLQADARSLLSBIQALH-AQNNGRKITLKREQSEK-----PSQELLEYNIQKQSQ 3103

QY 348 LGEHTRSIGPRESAKERYORSLE-----KELFF 378
DB 3104 MLEMQVELSSMKDRATELOEQLSSEKVMVAELKSELQAOTKLETTLKAQKHLEAF 3163

QY 379 AYDVFGIPFVDPDSWTPEEVIKRLQEKOKSRETAVRISQIEIGNLMKEITLVEETK 438
DB 3164 RLEV-----KDKTDEVHLINDTLASEQKKSRELOWALEKEKAKLGRS-----EERDK 3211

QY 439 SLDVSRLTREGGPL--LYEGISLTMSKLLN-----GYQVVM-DGVISDHECCOELQ 487
DB 3212 ELEDLKFSLESQKORNQLNLLLEQQKQLNLESQKTESQRMLYDAQLSQEGGRNLEQV 3271

QY 488 -----RLTNVAAT-----SGDGYRGQTSPTHTPNEKF----- 513
DB 3272 LLESEKVIAREMSLTDRERELHAQLQSSDG-TQCSRPLPSEDLKELQKLEKHSRI 3330

QY 514 -----XGTVTF-----KALKGOBKVPLQ-SAHLYNVTEKVRIM 549  
DB 3331 VELLNETEKYKLSLQTRQOMKDRQVHRTKLTQEQANTEGOKKHQSKVEDLQRL 3390  
QY 550 E----SYFLDTPLYFSYSHLVCRTEAEVQAEKDDSHPHVDNCILNAETLVCVKEPP 605  
DB 3391 EEKQVYKLDLEQ-RLQIMQEFQKQLEERESRRILYQ-----LNEPT 3439  
QY 606 AYTF-----BDYSAILVNGDFGNGFYFTLDAKTVAEVQCCGRAVFSFGTENPHG 660  
DB 3440 TWSLTSRTNRVWLOQKIEGETESN-----AKUEMN-----GGTGCNHE 3482  
QY 661 KVAVTRQRCALALWFTLDRHSEDRVQADDLVKMLFSPPEMD--LSQEQPLDAQQGP 718  
DB 3483 LEMIRQKLCQVASKQLVLPQKASERLQFETADDEDFWQENIDEIILQLKLTGQG-E 3541  
QY 719 BPAQESLSGS 728  
DB 3542 EPSLVSPSTS 3551

## RESULT 7

US-07-920-281C-2  
; Sequence 2, Application US/07920281C  
; Patent No. 5739026  
; GENERAL INFORMATION:  
; APPLICANT: Garoff, Henrik  
; APPLICANT: Liljestrom, Peter  
; TITLE OF INVENTION: DNA Expression Systems Based on  
; TITLE OF INVENTION: Alphaviruses  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/920,281C  
; FILING DATE: 13-AUG-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr., Gerald M.  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 828-103P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-241-1300  
; TELEFAX: 703-241-2848  
; TELEX: 248345  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2431 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-920-281C-2

Query Match 2.9%; Score 111; DB 1; Length 2431;  
Best Local Similarity 20.9%; Pred. No. 0.85;  
Matches 144; Conservative 68; Mismatches 253; Indels 224; Gaps 32;  
QY 75 CRTQCAADFWEFLDPWSPSPAQSG-----AGALRDLSPFGGLLRRAACLRCLGPP 127  
DB 83 CPNRSABD-ERLD-SYAKKLAASGKVLDRITAGKITD-----LQTWATP 127  
QY 128 AAHSLSEMELEFRKSPYNYLQVAFKINKLEKAVAAAHFTFVGNPEHMEWQNLDDYQ 187

DB 128 DAESEPTFLCHTDTVCRTA-----AEVAVY-----QDVYAVHA-----PTSL-----YHQ 166  
QY 188 TMSGVKEADFKDLETPHMQE-----  
DB 167 AMKGVRTAIWTFGDTTTFPMFDALAGAYTTATNWADEQV-QARNIGLCAASLTEGRGLK 226  
QY 209 -----FRLGVELYSEEQPOEAVPHLEAALQEFYFVAYBECALCEGPDYD 253  
DB 227 SILRKKQLKPCDVTWFSVGSGLYTESRKLRLSWHLPSVFLHKGQSFTCR--CDTIVSCE 284  
QY 254 GNYLLE-----YNADLFOAITDHYIQLVNLCKQNCVTELASHPSRKPDEFLPSH--- 303  
DB 285 GYVVKKITMCPGLYKTVGYAVTYHAEGLVCK--TTDTVKGERVSFPVCTVVPSTICD 341  
QY 304 -----YNYLQFAYVNYGNYTQAGECAKTYLLFFPNDENVNQNLA-Y 343  
DB 342 QMTGILATDVTPEDAQKLLVGLNQRIVN-GRTOENTMTMKVYL--PIVAVAFSKWRE 398  
QY 344 YAAMLGEHTRSIGPRESA-----KEYRQSRLEKELLFPAYDVFGIP-----FVD 389  
DB 399 YKADLDD--KPLGVRRERSLTCCCLWAFKTRKMTMYKKP---DTQTVIKVPSSEFNSFVI 453  
QY 390 PDSWTPPEEVIKPR-----LOEKOKSRETAVRISQEIGNLMKEITLVEBKTESL--- 440  
DB 454 PSLWSTGLAIPVRSRIKMLAKTKRELI PVIDASSARDAEQEERLEAELTREALPPL 513  
QY 441 -----DVSRLTREGGPLLYE-----GISLTMSKLLNGYQVRVMDGVISDH 481  
DB 514 VPIAPAEVGVVDVDEVELEYHAGAGVETPRSAKLVTAQPNVDLLGNVVLSPQTVLKSS 573  
QY 482 E-----CQELQRLTNVAATSG---DGYRGQTSFPHTPNEKFYGVTVFKALKGOEGKVP 531  
DB 574 KLAPVHPLAEQVKIITHNGRAGGVQVDGVRV--LPCGSAIPVPEFQALS----- 623  
QY 532 LQSAHLYNVNTEKVRIMESYFRDTPLYFSYSHLVCRTEAE-----VOAERKDDSHPVH 587  
DB 624 -ESATVYVNEREFVNR-----KLYHIAVHGPSLNTDEENYKVRARTDAEVVD 672  
QY 588 VDN--CILNAET--LVCVKE---PPAYTF 609  
DB 673 VDKKCCVKEEASGLVLVGLTNPFPHEF 701

## RESULT 8

US-08-466-277-2  
; Sequence 2, Application US/08466277  
; Patent No. 6190666  
; GENERAL INFORMATION:  
; APPLICANT: Garoff, Henrik  
; APPLICANT: Liljestrom, Peter  
; TITLE OF INVENTION: DNA Expression Systems Based on  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,277  
; FILING DATE: 06-Jun-1995  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/920,281  
; FILING DATE: <Unknown>



Db 365 YHVDYSN-TQSSGNISRGSSNPIDLNPIIISTCIRNSFYKAIAGSSVLVNFKGTQG 423  
Qy 529 ----KVPLQSA-----HLYNVTEKVRIMESYFRDTPLYFSYHL 566  
Db 424 YAFQAPTGGAWDHSFIESDGAPEGHKLNITYTSPGDTLRDFINVTLISTP----- 475  
Qy 567 VCRFAIEVQAE-----KDDSHPVHVDNCILNAETLVC- 600  
Db 476 ----TINELSTEKIKGPAEKYIKNOGIMKYKPEYINGAQPVNLE-----NQOTLIFE 527  
Qy 601 --VKEPAYTRDYSAILYNGDFDGNFYFTELDKATVTAEVQPCGAVGSSGT--E 656  
Db 528 PHASKTAQYTIR-----IRYASTQGTGKYFRDLNQELQTLNPTSHN-----GYVTGNIGE 578  
Qy 657 N-----PHGVKAVTRGQRCALALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEOPLD 712  
Db 579 NYDLTYTIGSYTITEGNH-----TLQIQHNDKNGMVLDD---RIEFVFK---DSLQDSP-- 624  
Qy 713 AQQPPPEPAQBSLSGSSEKSP 732  
Db 625 -QDSPPEVHSTIIFDKSSP 643

## RESULT 10

US-08-158-232-6  
; Sequence 6, Application US/08158232  
; Patent No. 5596071  
; GENERAL INFORMATION:  
; APPLICANT: Payne, Jewel  
; APPLICANT: Kennedy, M. Keith  
; APPLICANT: Randall, John Brooks  
; APPLICANT: Meier, Henry  
; APPLICANT: Uick, Heidi Jane  
; APPLICANT: Foncetrada, Luis  
; APPLICANT: Schnepf, H. Ernest  
; APPLICANT: Schwab, George E.  
; APPLICANT: Fu, Jenny  
; TITLE OF INVENTION: No. 5596071el Bacillus thuringiensis Toxins Active  
; TITLE OF INVENTION: Against Hymenopteran Pests  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David R. Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/158,232  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/887,980  
; FILING DATE: 22-MAY-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/797,645  
; FILING DATE: 25-NOV-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/703,977  
; FILING DATE: 22-MAY-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: M/SCU104.C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800

; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1257 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Bacillus thuringiensis  
; INDIVIDUAL ISOLATE: PS33F2  
; IMMEDIATE SOURCE:  
; CLONE: E. coli NM522 (pMYC2316) B-18785  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..1257  
; US-08-158-232-6  
  
Query Match 2.9%; Score 110.5; DB 1; Length 1257;  
Best Local Similarity 18.2%; Pred. No. 0.3;  
Matches 113; Conservative 90; Mismatches 218; Indels 199; Gaps 31;  
  
Qy 222 QEAVPHLEAALQEFVAYEECRALCEGPDYDGYNLEYNADLFQAITDH--YIQVLNCK 279  
Db 114 BELKQIEALIQDITNYQD--AINQKKFD-----SLQKTINLYTVAIDNNYDTAKTL 166  
Qy 280 QNCVTELASHPSREKPFDFLPSHYNYLQFAYINIGNYTOAGECAKTYLLFFPNDVMNQ 339  
Db 167 ENLNSILTSDIS-----IFIPEGVETGGLPYAM-----VANAHILLRLDAIVNAE 212  
Qy 340 NLAY-----YAMLGEEHTRSI-----GPRESAKEYQRSLEKELFFA 379  
Db 213 KLGFSDKEVDTHKKYIKMTIHNTHEAVIKAFNLGLDKFKSLDVNSYNKKKANYIKGMTVM 272  
Qy 380 YDVFGI-PFVDPDGSWTPE-----EVIKRLQEKQKSERETAVRISQEIENL 424  
Db 273 LDVALWTFDPDHYQKEVEIEFTRTISSPIYQPVKKNQNTSSIVPSDLFHYQ--GDL 330  
Qy 425 MK-EIETLVEBKTESLDVSRLTREGGPLLYEGISLTWNKSLNGYORVMDGVISDHEC 483  
Db 331 VKLEFST-----RTDNDGLAKIFTGIRNTFYKS-----PNTHT 364  
Qy 484 QELQLRTNVAATSGDVGSGTSP-----HTP-----NEKFY-----GVTVFKALKGOEG 528  
Db 365 YHVDYSN-TQSSGNISRGSSNPIDLNPIIISTCIRNSFYKAIAGSSVLVNFKGTQG 423  
Qy 529 ----KVPLQSA-----HLYNVTEKVRIMESYFRDTPLYFSYHL 566  
Db 424 YAFQAPTGGAWDHSFIESDGAPEGHKLNITYTSPGDTLRDFINVTLISTP----- 475  
Qy 567 VCRFAIEVQAE-----KDDSHPVHVDNCILNAETLVC- 600  
Db 476 ----TINELSTEKIKGPAEKYIKNOGIMKYKPEYINGAQPVNLE-----NQOTLIFE 527  
Qy 601 --VKEPAYTRDYSAILYNGDFDGNFYFTELDKATVTAEVQPCGAVGSSGT--E 656  
Db 528 PHASKTAQYTIR-----IRYASTQGTGKYFRDLNQELQTLNPTSHN-----GYVTGNIGE 578  
Qy 657 N-----PHGVKAVTRGQRCALALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEOPLD 712  
Db 579 NYDLTYTIGSYTITEGNH-----TLQIQHNDKNGMVLDD---RIEFVFK---DSLQDSP-- 624  
Qy 713 AQQPPPEPAQBSLSGSSEKSP 732  
Db 625 -QDSPPEVHSTIIFDKSSP 643  
  
RESULT 11  
US-08-304-626-6  
; Sequence 6, Application US/08304626  
; Patent No. 5616495  
; GENERAL INFORMATION:



APPLICANT: Payne, Jewel M.  
APPLICANT: Kennedy, M. Keith  
APPLICANT: Randall, John Brooks  
APPLICANT: Meier, Henry  
APPLICANT: Uick, Heidi Jane  
APPLICANT: Focerrada, Luis  
APPLICANT: Schnepf, Harry E.  
APPLICANT: Schwab, George E.  
TITLE OF INVENTION: No. 5616495el Bacillus thuringiensis Isolates  
TITLE OF INVENTION: Active Against Hymenopteran Pests and Genes Encoding  
TITLE OF INVENTION: Hymenopteran-Active Toxins  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/304,626  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/897,980  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/SCJ 104  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1257 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
INDIVIDUAL ISOLATE: PS33F2  
CLONE: E. coli NM522(pMYC2316) B-18785  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..1257  
US-08-304-626--6  
Query Match 2.9%; Score 110.5; DB 1; Length 1257;  
Best Local Similarity 18.2%; Pred. No. 0.3;  
Matches 113; Conservative 90; Mismatches 218; Indels 199; Gaps 31;  
QY 222 QEAVPHLEALQSYFVAYEACRALCEGPDYDGYNLYENADLFOAITH--YIQVLNCK 279  
DB 114 ELKPOELAIQDDITNYQD--AINOKKFD-----SLQKTINLYTVAINDDYVTKATQL 166  
QY 280 QNCVTLEASHPSREKPFEDPLPSHYNYLQPAYYNIQNTQAGCAKTYLLFFPNDEVMNQ 339  
DB 167 ENLSILTSIS-----IFIPGYETGGLPYAM-----VANAHILLRDAIVNAE 212  
QY 340 NLAY-----YAMLGEETRST-----GPRESAKYRQSRSLLEKELLFPA 379  
DB 213 KLGFSDKEVDTHKKYIKMTIHNTHEAVIKAFNLGLDKFKSLDVNSYNNKXANIKGMTVMV 272

QY 380 YDVFGLPFVDPDSWTPB-----EVIKLEOKKOKSERETAVRISQEIGNL 424  
DB 273 LDLVALMPTDFPDHYQKEVEIEFTRTISSPIYQVPKMNQMTSSSVFSDLFHYQ--GDL 330  
QY 425 MK-BIETLVEBKTESLDVSRLTREGGPLLLEGISLTWNKLLNGYQYRVVMDGVISDHEC 483  
DB 331 VKLEFST-----RTDNDGLAKIPTGIRNTFYKS-----PNTHET 364  
QY 484 QELQRLTNVAATSGDVGRTQSP-----HTP-----NEKFY-----GYTVFKALKQOEG 528  
DB 365 YHVDPSYN-TOSSNGISRGSSNPIDLNPNPIISTCIRNSFYKAIGSSVLVNFADSTQG 423  
QY 529 ---KVLQSA-----HLVYVNTKVRIMESYFRDLTPLYFSYSHL 566  
DB 424 YAFQAPTGGAWDHSFIESDGAPEGHKLNYYTSPGDTLRDFINVYTLISTP----- 475  
QY 567 VCRFAIEVQAE-----KDDSHPHVDVNCILNATLVC- 600  
DB 476 ---TINELSTEKINGFPAEXYIKNOGIMKYKPEYINGAQPVNLE---NQOTLIFE 527  
QY 601 --VKEPPAYTRDYSAIILYLANGDFDGGNFYFTELDKATVTAEVQPCQGRAVGFSSGT--E 656  
DB 528 PHASKTAQYTR---TRYASTQGTGKGYFRLDNLOELQTLNPTSHN-----GYVTGNIGE 578  
QY 657 N----PHGKAVTRGQRCAIALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEQPLD 712  
DB 579 NVDLTYIGSYTITEGNH-----TLQIHNKNGWLD---RIEFVVK--DSLQDSP-- 624  
QY 713 AQQPPPEPAQESLSGSSEK 732  
DB 625 -QDSPEVHESTIIIFDKSSP 643

RESULT 12  
US-08-315-301A-6  
Sequence 6, Application US/08316301A  
Patent No. 5753492  
GENERAL INFORMATION:  
APPLICANT: Schnepf, Harry E.  
APPLICANT: Schwab, George E.  
APPLICANT: Payne, Jewel M.  
APPLICANT: Harva, Kenneth E.  
APPLICANT: Focerrada, Luis  
TITLE OF INVENTION: No. 5753492el Nematode-Active Toxins and Genes  
TITLE OF INVENTION: Which Code Therefor  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Saliwanchik & Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/316,301A  
FILING DATE: 30-SEP-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/871,510  
FILING DATE: 23-APR-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/693,018  
FILING DATE: 03-MAY-1991  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/565,544  
FILING DATE: 10-AUG-1990

Db	528	PHASKTAQYTIIR----	IRYASTQGTGKGYFRDLNQLNIFTSN-----GYVTGNICE	578
Qy	657	N----	PHGVKAVTRGQRCALWFTLDRPHSRDRVQADDLVKMLFSPSEMDLSQEQPLD	711
Db	579	NYDLTYTIGSYTITEGNE-----	TLQIQHNDKNGMVLDD--RIEFVPK--DSLQDSP--	624
Qy	713	AQGPBPAPQESLSGSSEKX	732	
Db	625	-QDSPPEVHESHTIIFDKSSP	643	
RESULT 13				
US-08-611-928-6				
; Sequence 6: Application US/08611928				
; Patent No. 5824792				
; GENERAL INFORMATION:				
; APPLICANT: Payne, Jewel				
; APPLICANT: Kennedy, M. Keith				
; APPLICANT: Randall, John Brooks				
; APPLICANT: Meier, Henry				
; APPLICANT: Uick, Heidi Jane				
; APPLICANT: Fonceerrada, Luis				
; APPLICANT: Schnepf, H. Ernst				
; APPLICANT: Schwab, George B.				
; APPLICANT: Fu, Jenny				
; TITLE OF INVENTION: No. 5824792el Bacillus thuringiensis Toxins Active				
; TITLE OF INVENTION: Against Hymenopteran Pests				
; NUMBER OF SEQUENCES: 51				
; CORRESPONDENCE ADDRESS:				
; ADDRESSEE: David R. Saliwanchik				
; STREET: 2421 N.W. 41st Street, Suite A-1				
; CITY: Gainesville				
; STATE: FL				
; COUNTRY: USA				
; ZIP: 32606				
; COMPUTER READABLE FORM:				
; MEDIUM TYPE: Floppy disk				
; COMPUTER: IBM PC compatible				
; OPERATING SYSTEM: PC-DOS/MS-DOS				
; SOFTWARE: PatentIn Release #1.0, Version #1.25				
; CURRENT APPLICATION DATA:				
; APPLICATION NUMBER: US/08/611,928				
; FILING DATE: 06-MAR-1996				
; CLASSIFICATION: 530				
; PRIOR APPLICATION DATA:				
; APPLICATION NUMBER: US 08/158,232				
; FILING DATE: 24-NOV-1993				
; APPLICATION NUMBER: US 07/887,980				
; FILING DATE: 22-MAY-1992				
; CLASSIFICATION: 530				
; PRIOR APPLICATION DATA:				
; APPLICATION NUMBER: US 07/797,645				
; FILING DATE: 25-NOV-1991				
; CLASSIFICATION: 530				
; PRIOR APPLICATION DATA:				
; APPLICATION NUMBER: US 07/703,977				
; FILING DATE: 22-MAY-1991				
; CLASSIFICATION: 530				
; ATTORNEY/AGENT INFORMATION:				
; NAME: Saliwanchik, David R.				
; REGISTRATION NUMBER: 31,794				
; REFERENCE/DOCKET NUMBER: M/SCJ104.C1				
; TELEPHONE: 904-375-8100				
; TELEFAX: 904-372-5800				
; INFORMATION FOR SEQ ID NO: 6:				
; SEQUENCE CHARACTERISTICS:				
; LENGTH: 1257 amino acids				
; TYPE: amino acid				
; STRANDEDNESS: single				
; TOPOLOGY: linear				
; MOLECULE TYPE: protein				
; HYPOTHETICAL: YES				

ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
INDIVIDUAL ISOLATE: PS33F2  
IMMEDIATE SOURCE:  
CLONE: E. coli NM522(pMYC2316) B-18785  
NAME/KEY: Protein  
LOCATION: 1..1257  
US-08-611-928-6

Query Match 2.9%; Score 110.5; DB 2; Length 1257;  
Best Local Similarity 18.2%; Pred. No. 0.3; Indels 199; Gaps 31;  
Matches 113; Conservative 90; Mismatches 218; Indels 199; Gaps 31;

QY 222 QEAVPHLEALQRYFYVAYEECRALCEGPDYDGYNYLEYNADLFOAITH--YIQVLNCK 279  
DB 114 BELKPOEALIQDITNYQD--AINQKFPD-----SLQKTNILYTVADNNDYVTAKTQL 166  
QY 280 QNCVTELASHPSREKPFEDFLPSHYNYLQFAYYIGNVTOAGECAKTYLLFFPNDEVMMQ 339  
DB 167 ENLSILTSDIS-----IFPEGETGGLPYAM-----VANAHILLRLDAIVNAE 212  
QY 340 NLAY-----YAAMLGEEHTRSI-----GPRSAKEYRORSLEKELLPFA 379  
DB 213 KLGFSDKVDTHKKYIKMTIHNTHEAVIKAFNLGLDKFKSLDVNSYKANKYIKGTEMV 272

222 QEAVPHLEALQRYFYVAYEECRALCEGPDYDGYNYLEYNADLFOAITH--YIQVLNCK 279  
114 BELKPOEALIQDITNYQD--AINQKFPD-----SLQKTNILYTVADNNDYVTAKTQL 166  
280 QNCVTELASHPSREKPFEDFLPSHYNYLQFAYYIGNVTOAGECAKTYLLFFPNDEVMMQ 339  
167 ENLSILTSDIS-----IFPEGETGGLPYAM-----VANAHILLRLDAIVNAE 212  
340 NLAY-----YAAMLGEEHTRSI-----GPRSAKEYRORSLEKELLPFA 379  
213 KLGFSDKVDTHKKYIKMTIHNTHEAVIKAFNLGLDKFKSLDVNSYKANKYIKGTEMV 272  
380 YDVFGL-PFYDPSWTPE-----EVIKRLQEKQKQSRFATVARSIGL 424  
273 LDVLALMPTDPHYQKEVIEFTRTSSPIQYVPEKMNQNTSSSVPSDLFHYQ--GDL 330  
425 MK-EIEFLVEEKYSLDVSRLTRREGGPLLYEGISLTMSKLLNGYQVVMGDVSDHEC 483  
331 VKLEFST-----RTUNDGLAKIFTGIRNTFYKS-----PNTHET 364  
484 QEQRLTNVAATSGDYGRTGTP-----HTP-----NEKFY-----GVTVFALKLQGG 528  
365 YHVDPSN--TQSSGNTSRGSSNPIDLNPIISTCIRNSFYKAIAGSSVLYNFKDGTQ 423  
529 -----KVPLOQA-----HLYNVTQKVRIMESFRDLTPYFSYHL 566  
424 YAFQAAPTGGAWDHSFESDGAPEGHKLNVIYTSPTDLRDFINVTLLSTP----- 475  
567 VCRTAIEVQAE-----KDDSHPVHVDNCLNAETLVC- 600  
476 -----TINELTEKLGPPAEKVIKQIMKYKPEYINGAQPVNLE-----NQOTLIE 527  
601 --VKEPPAYTRDYSAIYLNGLDGGNFYFTTDLAKTVTAEVQPCGAVGFSST--E 656  
528 FHASKTAQYTIIR-----IRVASTQGTGKGYRLDNQELQTLNIPTSN-----GYVTGNIGE 578  
657 N-----PHGVKAVTGQRCALMFTLDPHRSERDRVQADDLVKMLFSPREMDSLQEQPLD 712  
579 NYDLTYTGSYTTIEGNH-----TLQIQRNDXNGMWLD-----RIEFVPR--DSLQDSP-- 624  
713 AQQGPPEPAOESLSGSESKP 732  
625 -QDSPEVHESITIFDKSSP 643

RESULT 14  
US-09-173-891-6  
Sequence 6, Application US/09173891  
Patent No. 6077937  
GENERAL INFORMATION:  
APPLICANT: Payne, Jewel  
APPLICANT: Kennedy, M. Keith  
APPLICANT: Randall, John Brooks  
APPLICANT: Meier, Henry  
APPLICANT: Uick, Heidi Jane  
APPLICANT: Foncerrada, Luis  
APPLICANT: Schnepf, H. Ernest  
APPLICANT: Schwab, George E.

APPLICANT: Fu, Jenny  
TITLE OF INVENTION: NO. 6077937el Bacillus thuringiensis Toxins Active  
TITLE OF INVENTION: Against Hymenopteran Pests  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/173,891  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/158,232  
FILING DATE:  
APPLICATION NUMBER: US 07/887,980  
FILING DATE: 22-MAY-1992  
APPLICATION DATA:  
PRIOR APPLICATION NUMBER: US 07/797,645  
FILING DATE: 25-NOV-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/703,977  
FILING DATE: 22-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: M/SCU104.C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1257 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bacillus thuringiensis  
INDIVIDUAL ISOLATE: PS33F2  
IMMEDIATE SOURCE:  
CLONE: E. coli NM522(pMYC2316) B-18785  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..1257  
US-09-173-891-6

Query Match 2.9%; Score 110.5; DB 3; Length 1257;  
Best Local Similarity 18.2%; Pred. No. 0.3;  
Matches 113; Conservative 90; Mismatches 218; Indels 199; Gaps 31;

QY 222 QEAVPHLEALQRYFYVAYEECRALCEGPDYDGYNYLEYNADLFOAITH--YIQVLNCK 279  
DB 114 BELKPOEALIQDITNYQD--AINQKFPD-----SLQKTNILYTVADNNDYVTAKTQL 166  
QY 280 QNCVTELASHPSREKPFEDFLPSHYNYLQFAYYIGNVTOAGECAKTYLLFFPNDEVMMQ 339  
DB 167 ENLSILTSDIS-----IFPEGETGGLPYAM-----VANAHILLRLDAIVNAE 212  
QY 340 NLAY-----YAAMLGEEHTRSI-----GPRSAKEYRORSLEKELLPFA 379  
DB 213 KLGFSDKVDTHKKYIKMTIHNTHEAVIKAFNLGLDKFKSLDVNSYKANKYIKGTEMV 272

```
QY 380 YDVGFI-PFVDPDSWTPE-----EVIKRLQKQKSERETAVRISQEIQNL 424
Db 273 LDVALWPTFDPDHYQKEVEIEFTRTISSPIYQVPVKMONTSSIVPSDLFHYQ--GDL 330
QY 425 MK-EIETLVEEKTESLDVSRLTREGGPELLYEGISLTWNSKLLNGYQVRVMDGVISDHEC 483
Db 331 VKLEFST-----RTDNDGLAKIFTGIRNTFYKS-----PNTHET 364
QY 484 QELQRLTNVAATSGDGYRGQTS-----HTP-----NEKFY-----GVTVFKALKLQGEQ 528
Db 365 YHVDPSYN-TQSSGNISRGSSNPIDLNPIIISTCIRNSFYKAIAGSSVLVNFKDGTOG 423
QY 529 ----KVPLQSA-----HLYYVTEKVRIMESYFRLDTPLYFSYSHL 566
Db 424 YAFQAAPTGGAWDHSFIESDGAPEGHKLNITYTSPGDTLRDFINVTYLTISTP----- 475
QY 567 VCRTAIEEVQAE-----KDDSHPVHVDNCILNAETLVC- 600
Db 476 ----TINELSTEKIKGPPAEKGYIKQGMKYYKPEYINGAQPVNLE-----NQOTLIFE 527
QY 601 --VKEPPAYTRDYSAIYLNGDFDGNFYFTELDAKTVTAEVQPCQGRAVGFSSGT--E 656
Db 528 PHASKTAQYTIIR-----IRYASTQGTGKYFRLDNQELQTLNPTSHN-----GYVTGNIGE 578
QY 657 N----PHGVKAVTRQORCAIALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEQPLD 712
Db 579 NYDLYTIGSYTITEGNH-----TLQIQHNDKNGMWLD---RIEFVPK--DSLQDSP-- 624
QY 713 AQQGPPPEPAQSLSGSESKP 732
Db 625 -QDSPPPEVHSTIIFDKSSP 643
```

## RESULT 15

```
US-09-076-137-6
; Sequence 6, Application US/09076137B
; Patent No. 6166195
; GENERAL INFORMATION:
; APPLICANT: Schnepf, Harry E.
; APPLICANT: Schwab, George E.
; APPLICANT: Payne, Jewel M.
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Poncerrada, Luis
; TITLE OF INVENTION: No. 6166195el Nematode-Active Toxins and Genes Which Code
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: MA-20CCD2
; CURRENT APPLICATION NUMBER: US/09/076,137B
; CURRENT FILING DATE: 1998-05-12
; EARLIER APPLICATION NUMBER: 08/316,301
; EARLIER FILING DATE: 1994-09-30
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1257
; TYPE: PRK
; ORGANISM: Bacillus thuringiensis
US-09-076-137-6
```

```
Query Match 2.9%; Score 110.5; DB 3; Length 1257;
Best Local Similarity 18.2%; Pred. No. 0.3;
Matches 113; Conservative 90; Mismatches 218; Indels 199; Gaps 31;

QY 222 QEAVPHLEALQEFYVAYEBCRALCEGFDYDGYNYLEVNADLFQAITH--YIQVLNCK 279
Db 114 EELKPOIEALIQODITNYQD--AINQKFD-----SLQKTINLYTVAIDNNDYVTAQTOL 166
QY 280 QNCVTELASHPSREKPPEDFLPSHYNYLQFAYYIGNYQAGSCAKTYLLFFPNDVEMNQ 339
Db 167 ENLSILTSIS-----IPIPEGYETGGLPYYAM-----VANAHILLERDAIVNAE 212
QY 340 NLAY-----YAAMLGEHTRSI-----GPRESAKEYRQSRSLLEKELLFFA 379
Db 213 KLGFSKDEVDTHKKYIKMTIHNHTEAVIKAFNLGLDKFKSLDVNSYNNKANYIKGMTEMV 272
```

Search completed: July 18, 2004, 09:45:17  
Job time : 60 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 09:58:44 ; Search time 978 Seconds  
(without alignments)  
3197.008 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 3870  
Sequence: 1 MAVRALKLLTLLAVVAAS.....PPEPAQESLSGSSEKPKDEL 736

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODELS=frame+ p2n.model -DBV=xlp  
-Q/cgr2\_1/USPTO\_spool\_p/US10045815/runat\_14072004\_123007\_16812/app\_query.fasta\_1.903  
-DBN\_Geneseq\_29Jan04 -QFMT=fastap -SUFFIX=std.rng -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USR=US10045815 @CGN 1 1 885 @runat\_14072004\_123007\_16812 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq\_29Jan04:\*  
1: Geneseq1980s:\*  
2: Geneseq1990s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001as:\*  
5: Geneseq2001bs:\*  
6: Geneseq2002s:\*  
7: Geneseq2003as:\*  
8: Geneseq2003bs:\*  
9: Geneseq2003cs:\*  
10: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3870	100.0	2600	3 AAC64725	Aac64725 Human tum
2	3861	99.8	2242	6 ABZ11346	Abz11346 Human pol
3	3846	99.4	2211	6 AAS17573	Aas17573 DNA encod
4	3846	99.4	2563	5 AAF93800	Aaf93800 Human cDN
5	3846	99.4	2585	4 AAH14589	Aah14589 Human cDN
6	3844	99.3	2524	4 AAH14513	Aah14513 Human cDN
7	3820	98.7	2577	4 AAD08489	Aad08489 Human sec
8	3801	98.2	2753	6 ABK94910	Abk94910 Human nov

9	3801	98.2	2753	6 ABK94933	Abk94933 Human nov
10	3792	98.0	2615	4 AAD08507	Aad08507 Human sec
11	3780	97.7	2829	3 AAC64724	Aac64724 Human tum
12	3726.5	96.3	2993	4 AAH16288	Aah16288 Human cDN
13	3645	94.2	2152	6 ABZ11345	Abz11345 Human pol
14	3572	92.3	2127	6 AAS17572	Aas17572 DNA encod
15	3323	85.9	2416	3 AAC64726	Aac64726 Mouse tum
16	3120	80.6	2322	3 AAC64727	Aac64727 Mouse tum
17	2148	55.5	1607	6 ABQ76716	Abq76716 Human leu
18	2147	55.5	1309	6 ABZ11344	Abz11344 Human pol
19	1751	45.2	3396	7 ADA53353	Ada53353 Human cod
20	1747	45.1	2127	9 ACF79584	Acf79584 Human pro
21	1747	45.1	2261	9 ACF79585	Acf79585 Human pro
22	1735	44.8	1595	3 AAC76943	Aac76943 Human ORF
23	1715.5	44.3	2665	7 ACA03965	Aca03965 cDNA down
24	1575.5	40.7	2789	4 AAH14459	Aah14459 Human cDN
25	1473	38.1	2305	7 ABX70413	Abx70413 DNA encod
26	1247.5	32.2	2129	8 AAL61038	Aal61038 Human pro
27	1186	30.6	2248	7 ABX34670	Abx34670 Human mdd
28	1108	28.6	703	4 AAH06792	Aah06792 Human cDN
29	1089.5	28.2	774	4 AAH06885	Aah06885 Human cDN
30	1072	27.7	682	4 AAH196734	Aah196734 Human neu
31	1040	26.9	652	5 AAF93998	Aaf93998 Primer sp
32	1029.5	26.6	738	4 AAH06612	Aah06612 Human cDN
33	977	25.2	550	7 ACA57060	Aca57060 Human adi
34	906.5	23.4	849	8 ADB83188	Adb83188 Human cDN
35	836.5	21.6	1023	4 AAS27088	Aas27088 cDNA enco
36	836.5	21.6	1023	4 ABK43918	Abk43918 DNA encod
37	836.5	21.6	1023	9 ADB93266	Adb93266 Human cDN
38	807.5	20.9	793	9 ACF79586	Acf79586 Human pro
39	804	20.8	1849	4 ABK43627	Abk43627 DNA encod
40	790	20.4	681	5 AAS80583	Aas80583 DNA encod
41	710.5	18.4	685	9 ACF79589	Acf79589 Human pro
42	708.5	18.3	684	9 ACF79589	Acf79589 Human pro
43	674	17.4	800	4 AAH06448	Aah06448 Human cDN
44	674	17.4	800	9 ACF79587	Acf79587 Human pro
45	664.5	17.2	649	9 ACF79593	Acf79593 Human pro

ALIGNMENTS

RESULT 1  
AAC64725  
ID AAC64725 standard; cDNA; 2600 BP.

XX	AC	AAC64725;
XX	DT	27-FEB-2001 (first entry)
XX	DE	Human tumour suppressor Gros1-S encoding cDNA SEQ ID NO:3.
XX	XX	Tumour suppressor; Gros1-L; Gros1-S; cell proliferation; regulation;
XX	XX	cancer; cytostatic; gene therapy; ss.
XX	OS	Homo sapiens.
XX	PN	WO200005047-A1.
XX	XX	02-NOV-2000.
XX	PD	26-APR-2000; 2000WO-JP002731.
XX	PF	26-APR-1999; 99JP-00118806.
XX	PR	(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX	PA	Wadhwa R, Sugihara T, Yoshida A;
XX	PI	WPI: 2000-687340/67.
XX	DR	P-PSDB; AAB36392.
XX	DR	Full-length tumor suppressor genes encoding Gros1-L, Gros1-S and mouse
XX	PT	homologs participating in regulation of cell proliferation, useful in

PT development of preventives and remedies of cancer.

XX PS Claim 1; Page 70-77; 114pp; Japanese.

XX The present sequence encodes the human tumour suppressor designated Gros1  
CC -S. Gros1-L and Gros1-S have cytosolic activity and can be used in gene  
CC therapy. Gros1-L and Gros1-S genes are useful in the development of drugs  
CC used to treat and prevent cancer

XX SQ Sequence 2600 BP; 608 A; 757 C; 722 G; 513 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 0 Length: 2600  
Score: 3870.00 Matches: 736  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 3 Gaps: 0

US-10-045-815-4 (1-736) x AAC64725 (1-2600)

QY 1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaSer 20  
DB 52 ATGGCGGTACGGGTTGAAGCTCTGACACACTGCTGGCTGCGTGGCGCTGCTCC 111  
QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40  
DB 112 CAAGCCGAGGTCGAGTCCGAGGAGGATGGGCATGCTGACGCTGATCTGCTTCGCC 171  
QY 41 GluGlyThrAlaAlaValAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
DB 172 GAGGGGACCGCAGCTACCGCGCGGGAGCTGGCCGGGTGGTCTGAGCATGGAACGG 231  
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
DB 232 CGCGTGGCTCCGGGAGGCGCTCCGGCGCTTCGCTGGCTGGCGCAGTGTGCC 291  
QY 81 AlaAspPheProTrpGluLeuLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
DB 292 GCCGACTTCCGCGGAGCTGGACCCGAGCTGGTCCCGCAGCGCGGCCAGGCTCGGGC 351  
QY 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyClyLeuLeuArgAlaAlaCysLeu 120  
DB 352 GCCGGCGCTCCGCGACCTGAGCTTCTCGGGGCGCTTCGCTGGCTGGCTGCTGCTG 411  
QY 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
DB 412 CGCGCTGCTCGCGCGCGCGCGCGCTCGCTCAGCGAGAGATGGAGCTGGAGTTTC 471  
QY 141 ArgLysArgSerProTrpGlnThrMetSerGlyValAlaValAlaValPheLysLeuGlu 160  
DB 472 CGCAAGCGGAGCGCTCAACTACTCGAGTGGCTGCTCAAGATCAACAAGTTGGAG 531  
QY 161 LysAlaValAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180  
DB 532 AAAGCTGTGTCGAGCACACACCTTCTCGTGGGCAATCTCGAGCACATGGAAATGCAG 591  
QY 181 GlnAsnLeuAspTrpTrpGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
DB 592 CAGAACCTAGACTATTACCAACCATCTCGAGTGAAGAGCGCGACTTCAAGGATCTT 651  
QY 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTrpSerGluGluGln 220  
DB 652 GAGACTCAACCCCATATGCAAGATTTTCGAGTGGAGTGGCTCTACTCAGAGGAACAG 711  
QY 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluThrPheValAlaValGlu 240  
DB 712 CCACAGGAAGCTGTGGCCCACTAGAGCGCGCTGCAAGAAATACCTTTGGCGCTATGAG 771  
QY 241 GluCysArgAlaLeuCysGluGlyProTrpAspTrpAspGlyTrpAsnTrpLeuGluTrp 260  
DB 772 GAGTGGCGTGGCTCTCGGAGGCGGCTTATGACTACGATGGCTACAACTACCTTGGATC 831

QY 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTrpIleGlnValLeuAsnCysLysGln 280  
DB 832 AACGCTGACCTCTCCAGGCCATCACAGATCATTCATCCAGGCTCTCACTGTAAGCAG 891  
QY 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
DB 892 AACTGTGTACGGAGCTTGTCTCCACCAAGTCGAGAGAACCCCTTGAAGACTCTCTC 951  
QY 301 ProSerHisTrpAsnTrpLeuGlnPheAlaTrpTrpAsnIleGlyAsnTrpThrGlnAla 320  
DB 952 CCATCCCATTAATAATATCTGAGTTGCTTACTATACATTTGGGAATATTACCAAGCT 1011  
QY 321 GlyGluCysAlaLysThrTrpLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
DB 1012 GGTGAATGTGCCAAGACCTATCTCTCTTCTCCCAATGACGAGGTGATGAACCAAAAT 1071  
QY 341 LeuAlaTrpTrpAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
DB 1072 TTGGCTATTATGACGCTATCTTGGAGAAACACACACAGATCCATCGGCCCCCGTGAG 1131  
QY 361 SerAlaLysGluTrpArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaTrp 380  
DB 1132 AGTGCACAGGAGTACCGACAGCGAGCTACTGGAAAAGAACTGCTTTCTTCCTTAT 1191  
QY 381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400  
DB 1192 CATGTTTTTGGAAATCCCTTTGTGATCCGGATTCAGGACTCCAGAGAAAGTGAATCCC 1251  
QY 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
DB 1252 AAGATTTGCAAGAGAAACAGAGTCAAGACGGAACAGCCGTACGATCTCCAGGAG 1311  
QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
DB 1312 ATTTGGAACTTATGAGGAAATCGAGACCTTGTGGAAGAGAAGACCAAGGAGTCACTG 1371  
QY 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTrpGluGlyIleSerLeuThr 460  
DB 1372 GATGTGACGACACTGACCCGGGAAGGTGGCCCTCTGCTGTATGAAGGATCAGTCTCACC 1431  
QY 461 MetAsnSerLysLeuLeuAsnGlyTrpGlnArgValValMetAspGlyValIleSerAsp 480  
DB 1432 ATGAACTCCAACTCTCTGAAATGGTTACCGCGGTGGTGTGGACGCGTAACTCTGAC 1491  
QY 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTrp 500  
DB 1492 CACGAGTGTACGAGCTGACAGAGCTGACCAATGTGCAGCAACCTCAGGAGATGGGTAC 1551  
QY 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTrpGlyValThrValPheLys 520  
DB 1552 CGGGTTCAGACTCCCCACATCTCCCAATGAAAGTTCTATGGTGTCTGCTTCAAA 1611  
QY 521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTrpTrpAsn 540  
DB 1612 GCGCTCAAGCTGGGGCAAGAGGCAAGATTCCTCTGACAGTGGCCACCTGTACTACAAC 1671  
QY 541 ValThrGluLysValArgArgIleMetGluSerTrpPheArgLeuAspThrProLeuTrp 560  
DB 1672 GTGACGGAAGATGGCGCGCATCATGGATCTCTTCCGCTGGATACGCCCTCTAC 1731  
QY 561 PheSerTrpSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
DB 1732 TTTTCTCTACTCTCATCTGGTGTGGCGCACTGCCATCGAAGAGGTCCAGGAGAGGAAG 1791  
QY 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
DB 1792 GATGATGTCATCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAG 1851  
QY 601 ValLysGluProProAlaTrpThrPheArgAspTrpSerAlaIleLeuTrpLeuAsnGly 620  
DB 1852 GTCAAGAGGCCCCCAGCGCTACACCTTCCGCGACTACAGCGCCATCTTACTTAAATGGG 1911  
QY 621 AspPheAspGlyGlyAsnPheTrpThrGluLeuAspAlaLysThrValThrAlaGlu 640

Db 1912 GACTTCGATCGCGGAAACTTTTATTTCACTGAACCTGATGCCAAGACCGTGACGGCAGAG 1971  
 Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
 Db 1972 GTGACGCTCAGTGTGAAGAGCGGTGGGATTCTTTCAGGCACTGAAACCCACATGGA 2031  
 Qy 661 ValIleAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTTPheThrLeuAspPro 680  
 Db 2032 GTGAAGCTGTCCACAGGGGGCAGCGCTGTGCCATCGCCTGTGGTTACCTGGACCT 2091  
 Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValIleMetLeuPheSerPro 700  
 Db 2092 CGACACAGCGAGCGGACAGGGTGCAGGCAGATGACCTGTGAGGATGCTTTCAGCCCA 2151  
 Qy 701 GluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
 Db 2152 GAGAGATGAGACCTCTCCAGGAGCAGCCCTGGATGCCAGCAGGGCCCCCGACCT 2211  
 Qy 721 AlaGlnGlnSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
 Db 2212 GCACAGAGTCTCTCAGGACGTGAATCGAAGCCCAAGGATGAGCTA 2259

## RESULT 2

ABZ11346  
 ID ABZ11346 standard; cDNA; 2242 BP.  
 XX  
 AC ABZ11346;  
 XX  
 DT 20-JAN-2003 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 228.  
 XX  
 KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
 KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
 KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
 KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
 KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;  
 KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
 KW haemostatic; vulnary; fungicide; antibacterial; virucide; protozoacide;  
 KW antiarthritic; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200270539-A2.  
 XX  
 PD 12-SEP-2002.  
 XX  
 PF 05-MAR-2002; 2002WO-US005095.  
 XX  
 PR 05-MAR-2001; 2001US-00799451.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PA Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
 PI Wehrman T, Wang J, Wang D, Dermanac RT;  
 XX  
 DR WPI; 2002-759812/82.  
 DR P-FSDB; ABP69129.  
 XX  
 PT New polynucleotides comprising sequences assembled from expressed  
 PT sequence tags (ESTs), useful for treating cell-proliferative,  
 PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet  
 PT or coagulation disorders.  
 XX  
 PS Claim 1; SEQ ID NO 228; 1012pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated polynucleotide (I) comprising a  
 CC nucleotide sequence selected from any of 948 sequences (ABZ11119-  
 CC ABZ12066) or their mature protein coding portion, active domain coding  
 CC protein or complementary sequences. The polynucleotides are useful for  
 CC identifying expressed genes or for physical mapping of human genome. The

CC encoded polypeptides (ABP6902-ABP69849) are useful as molecular weight  
 CC markers, as a food supplement, for generating antibodies, in medical  
 CC imaging, screening and diagnostic assays and for treating cell-  
 CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's  
 CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,  
 CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,  
 CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver  
 CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),  
 CC arthritis, etc. Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WFO at ftp.wipo.int/pub/published\_pct\_sequences

XX  
 SQ Sequence 2242 BP; 518 A; 655 C; 625 G; 444 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 0 Length: 2242  
 Score: 3861.00 Matches: 734  
 Percent Similarity: 99.86% Conservative: 1  
 Best Local Similarity: 99.73% Mismatches: 1  
 Query Match: 99.77% Indels: 0  
 DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x ABZ11346 (1-2242)

Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
 Db 32 ATGGCGGTACGGCGGTTGAAGCTGCTGACCACACTGCTGCTGCTGCTGCTGCCCTCC 91  
 Qy 21 GlnAlaGluValGluSerGluAlaGlyTTPGlyMetValThrProAspLeuLeuPheAla 40  
 Db 92 CAAGCCGAGGTGCGATCCGAGGAGGATGGGCGATGGTGACGCGCTGATCTCTTCGCC 151  
 Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTTPProGlyValValLeuSerMetGluArg 60  
 Db 152 GAGGGGACCGGAGCTACTCGCGGGGAGCTGGCCGGGGTGGTCTCTGAGCATGGAACGG 211  
 Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
 Db 212 GCGTTCGCTCCGGGAGCGCCCTCCGCGCCCTTCGCTGCGCTGCCGACCCAGTGTGCC 271  
 Qy 81 AlaAspPheProTTPGluLeuAspProAspTTPSerProSerProAlaGlnAlaSerGly 100  
 Db 272 GCGGACTTCCGCTGGGAGCTGGAACCCGACTGGTCCCGACGCGCGCCGCGCTCGGGC 331  
 Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLeu 120  
 Db 332 GCGCGCGCCCTCGCGGACCTGAGCTTCTTCGGGGGCGCTTCGCTGCGCTGCGCTGCC 391  
 Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe 140  
 Db 392 CGCGCTGCTCGGGCGCGCGCGCCACTCGCTCAGCGAAGAGATGGAGCTGGAGTTC 451  
 Qy 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
 Db 452 CGCAAGCGGAGCCCTACACACTACTCGAGGTGCGCTACTTCAAGATCAACAAGTTGGAG 511  
 Qy 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180  
 Db 512 AAAGCTGTTCGTCAGACACACACCTTCTTCGCTGGGCAATCTCTGAGCACATGGAATGAG 571  
 Qy 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
 Db 572 CAGAACCTTAGACTATTACCAACCATGCTGGAGTGAAGAGAGGCGCGACTTCAAGATCTT 631  
 Qy 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220  
 Db 632 GAGACTCAACCCCATATGCAAGAAATTCGACTGGAGTGGCACTCTACTCAGAGAACAG 691  
 Qy 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
 Db 692 CCACAGGAAGCTGTGCCCCACCTAGAGGCGCGCTGCAAGAATACTTTGTGCCCTATGAG 751  
 Qy 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260

Db 752 GAGTGGCGTGCCTCTCGAAGGCGCCCTATGACTACGATGGCTACACTACCTTGAGTAC 811  
Qy 261 AsnAlaSerPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
Db 812 AACGCTGACCTCTTCAGGCGCATCACAGATCATTCATCCAGGTCCTCACTGTAAGCAG 871  
Qy 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
Db 872 AACTGTGTACGAGCTGTGCTCCACCCCAAGTCGAGAGAAGCCCTTGAAGACTTCCTC 931  
Qy 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
Db 932 CCATCGCATATATAATTATCTGCAGTTTGCCTTACTATTAACATTTGGAAATATATACAAAGCT 991  
Qy 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
Db 992 GGTGAATGTGCCAAGACCTATCTCTCTTCCCAATCAGCAGGTGATGAACCAAAAT 1051  
Qy 341 LeuAlaTyrTyrAlaAlaMetLeuGluGluLysHisThrArgSerIleGlyProArgGlu 360  
Db 1052 TTGCGCTTATTATGAGCTATGCTTGGAGAGAACAACACAGATCCATCGGCCCCCGTGAG 1111  
Qy 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr 380  
Db 1112 AGTCCAAAGGATACCGACGCGAGCTACTGGAAAAGAACTGCTTTTCTTCGCTTAT 1171  
Qy 381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400  
Db 1172 GATGTTTTTGGAAATTCCTTTGTGGATCCGATTCATGGACTCCAGAAAGATGATCC 1231  
Qy 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
Db 1232 AAGAGATTGCNAGAAACACAGAGTCAGACGGGAAACACCGGTACGCATCTCCAGGAG 1291  
Qy 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
Db 1292 ATTGGGAACCTTATGAAGAAATCGAGACCTTGTGGAGAGAAGCAACCAAGGAGTCACTG 1351  
Qy 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuTyr 460  
Db 1352 GATGTGACGACACTGACCCGGAAGTGGCCCCCTGCTGTATGAAGCATCAGTCTCACC 1411  
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
Db 1412 ATGAATCCAAACTCTCTGAATGTTACCAGCGGTGTGTGATGACGCGTAACTCTGAC 1471  
Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaIleThrSerGlyAspGlyTyr 500  
Db 1472 CACGAGTGTACGAGAGCTGCGAGACTACCAATGTGGCAGCAACCTCAGGAGATGCTAC 1531  
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
Db 1532 CGGGGTGAGACCTCCCACTACTCCCAATGAAGAATTCATGGTGTCTGCTCTTCAAA 1591  
Qy 521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
Db 1592 GCCCTCAAGCTGGGGCAAGAGGCAAAAGTTCTCTGCGAGAGTCCCACTGTACTACAAC 1651  
Qy 541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
Db 1652 GTGACGAGAAAGTGGCGGCATCATGAGTCTTACTTCCGCTGTGATAGCCCTCTTAC 1711  
Qy 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArgLys 580  
Db 1712 TTTTCTTACTCTCATCTGGTGTGGCGCACTGCCATCGAAGAGGTCCAGGCGAGAGGAAG 1771  
Qy 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
Db 1772 GATGATAGTATCATCGATGCCAGTGGACACTGATCTCTGAATGCCAGAGACCTTCGTGT 1831  
Qy 601 ValLysGluProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620

Db 1832 GTCAAGAGAGCCCCCAGCCCTACACCTTCGCGACTACAGCGCATCTTTACTAAATGGG 1891  
Qy 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
Db 1892 GACTTCGATGCGGAAACTTTTATTTCACTGAATGGATGCCAAGACCGTGACGCGCAG 1951  
Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
Db 1952 GTGACGCTCATGTGTGGAAGAGCCGTGGATTTCTTTCAGCCTGAAACCCCATGGA 2011  
Qy 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680  
Db 2012 GTGAGGCTGTTCACACGAGGCGAGCGCTGTGCCATCGCCCTGTGTTTCCCTCGACCT 2071  
Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
Db 2072 CGACACAGCAGCGGACAGGGTGCAGGCAGATGACCTGGTGAAGATGCTCTTCAGGCCA 2131  
Qy 701 GluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
Db 2132 GAAGAGATGGACCTCTCCAGAGCAGCCCTGGATGCCCGCAGCGGCCCGCCGACCT 2191  
Qy 721 AlaGlnGluSerLeuSerGlySerGlySerLysProLysAspGluLeu 736  
Db 2192 GCACAAGAGTCTCTCTCAGGCAGTGAATCGAAGGCCAAGGATGAGCTA 2239  
RESULT 3  
AAS17573  
ID AAS17573 standard; cDNA; 2211 BP.  
XX  
AC AAS17573;  
DT 26-FEB-2002 (first entry)  
XX  
DE DNA encoding novel secreted protein #2.  
XX  
KW Secreted protein; cytostatic; immunosuppressive; vulnary; vaccine;  
KW antiinflammatory; neuroprotective; nephrotropic; cardiovascular; human;  
KW cancer; autoimmune disease; wound healing disorder; infection;  
KW haematopoietic disorder; inflammatory disorder; infertility;  
KW neurological disease; psychiatric disease; cardiovascular disease;  
KW respiratory disease; renal; gastrointestinal; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
CDS 1..2211  
FT /\*tag= a  
FT /product= "Human secreted protein"  
XX  
PN WO200179454-A1.  
XX  
PD 25-OCT-2001.  
XX  
PP 11-APR-2001; 2001WO-US011797.  
XX  
PR 13-APR-2000; 2000US-0196603P.  
PR 24-APR-2000; 2000US-0199417P.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
PI Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z;  
XX  
DR WPI; 2002-061975/08.  
DR P-PSDB; AAU09861.  
XX  
PT New secreted proteins or polypeptides, useful for treating e.g. cancer,  
PT autoimmune diseases, wound healing disorder, infections, hematopoietic  
PT disorders, inflammatory disorders, infertility, cancer.  
XX  
PS Claim 2; Page 35-36; 92pp; English.  
XX



CC The invention relates to an isolated novel secreted polypeptide (I) and  
CC polynucleotide (II). (I) and (II) are useful for treating cancer,  
CC autoimmune diseases, wound healing disorder, infections, haematopoietic  
CC disorders, inflammatory disorders, infertility, neurological and  
CC psychiatric diseases, cardiovascular diseases, respiratory diseases,  
CC renal diseases, or gastrointestinal diseases. These may also be used to  
CC treat diseases, abnormalities and disorders caused by abnormal  
CC expression, production, function and/or metabolism of the genes, as  
CC vaccines for inducing immunological response in a mammal, and in  
CC screening methods for detecting the effect of added compounds on the  
CC production of mRNA and polypeptide in cells. The polypeptides can be used  
CC as immunogens to produce antibodies immunospecific for the polypeptides,  
CC and to identify membrane-bound or soluble receptors. The polynucleotides  
CC may be used as diagnostic reagents, in chromosome localisation studies,  
CC and in tissue expression studies. The present sequence represents the  
CC coding sequence of novel human secreted protein #2  
XX  
SQ Sequence 2211 BP; 512 A; 649 C; 612 G; 438 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 0 Length: 2211  
Score: 3846.00 Matches: 733  
Percent Similarity: 99.59% Conservative: 0  
Best Local Similarity: 99.59% Mismatches: 3  
Query Match: 99.38% Indels: 0  
DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x AAS17573 (1-2211)

QY	1	MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValValAlaAlaAlaSer	20
DB	1	ATGGCCGTTACGCGGTTGAAGCTGTGACACACTGCTGGCTGTGCGCCGCTGCTCC	60
QY	21	GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla	40
DB	61	CAAGCCGAGTCGAGTCCGAGGCGAGGATGGGCGATGGTGACGCTGATCTGCTCTCGCC	120
QY	41	GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg	60
DB	121	GAGGAGACCGACCTACCGCGCGGGGACTGGCCCGGGGTGGTCTGAGCATGGAAAGG	180
QY	61	AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla	80
DB	181	GCCTCGGCTCCGGGCGAGCCCTCCGGCCCTTCGGCTCGGCTCGCGACCCAGTGTGCC	240
QY	81	AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly	100
DB	241	GCGACTTCCGTCGGGAGTGGACCCGCGACTGGTCCCGGCGCGCCGAGGCTCGGGC	300
QY	101	AlaGlyAlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgArgAlaAlaCysLeu	120
DB	301	GCCCGCCCTCGCGGACCTGAGCTTCTTCGGGGGCTTCTGGCTCGCGCTGCTGCTG	360
QY	121	ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe	140
DB	361	CGCCGCTGCTCGGCGCGCGCGCCGCTCGCTCAGCGAAGAGATGGAGTTC	420
QY	141	ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysLeuLeuGlu	160
DB	421	CGCAAGCGAGCCCTACCACTACCTGCGAGTGGCTGCTTCAAGATCAACAAGTTGGAG	480
QY	161	LysAlaValAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln	180
DB	481	AAAGCTGTTGCTGCAGACACACACTTCTCGTGGGCAATCCTGAGCACATGGAAATGCAG	540
QY	181	GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysLeu	200
DB	541	CAGAACCCTAGACTATTACCAAAACCATGTCTGGAGTGAAGGAGCCGCACTTCAAGATCTT	600
QY	201	GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGln	220
DB	601	GAGACTCAACCCCATATGCAAGNATTCGATGGGAGTGGGACTCTACTACAGAGGACAG	660

QY	221	ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu	240
DB	661	CCACAGGAGGTGTGTGCCCACTAGAGCGGGCTGCAAGATATCTTTGTGGCCCTATGAG	720
QY	241	GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr	260
DB	721	GAGTCCGCTGCTGTGGAAGGCGCTTATGACTACGATGGCTACCACTACCTTGAGTAC	780
QY	261	AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln	280
DB	781	AACGCTGACCTTCTCCAGGCCATCACAGATCATTCATCCAGGCTCTCACTGTAAAGAG	840
QY	281	AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu	300
DB	841	AACGTGTACGGAGCTTGTCTCCACCACCAAGTCAGAGAAAGCCCTTTGAAGACTTCCTC	900
QY	301	ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla	320
DB	901	CCATCGCAATTATAATATCTGCAGTTTGCCTACTATACATTTGGGAATATTATACAGGCT	960
QY	321	GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn	340
DB	961	GTTGAATGTGCCAAGACCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1020
QY	341	LeuAlaTyrTyrAlaAlaMetLeuGlyGluHisThrArgSerIleGlyProArgGlu	360
DB	1021	TTGGCTATTATGCGACTATCTTTGGAGAAACACACACAGATCCATCGGCCCTCGTGG	1080
QY	361	SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaTyr	380
DB	1081	AGTGCAAGGAGTACCGACGAGGCTACTTGAAAAGAACTGCTTTCTTCTGCTTAT	1140
QY	381	AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro	400
DB	1141	GATGTTTTTGGAAATCCCTTTGTGATCCGGATTCATGGACTCCAGAGAAAGTATTCCC	1200
QY	401	LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu	420
DB	1201	AAGAGATTGCAAGAAACACAGAGTCAGAACGGGAAACAGCCGACCGATCTCCAGAGG	1260
QY	421	IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu	440
DB	1261	ATTGGAACTTATGAAGAAATCGAGACCTCTTGGAGAGAGAACCAAGAGTCACTG	1320
QY	441	AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr	460
DB	1321	GATGTGACAGACTGACCCGGAAAGTGGCCCTCTGCTGTATGAAGCATCAGTCTCACC	1380
QY	461	MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp	480
DB	1381	ATGAACTCCAACTCTGATGGTTCACGCGGGTGTGATGGACGGCTAATCTCTGAC	1440
QY	481	HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr	500
DB	1441	CACGAGTGTACGAGGCTGCAGAGACTGACCAATCTGGCAGCAACCTCAGGAGATGGCTAC	1500
QY	501	ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys	520
DB	1501	CGGGTTCAGACTCCCACTACTCCCAATGAAAGTCTTATGTGTGCTGCTCTTCAAA	1560
QY	521	AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn	540
DB	1561	GCCCTCAAGCTGGGCAAGAGGCAAAAGTCTCTCTGAGAGTGGCCCTCTACTACAAC	1620
QY	541	ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr	560
DB	1621	GTGACGGAAGAGTGGCGCGCATCATGGAGTCTTCTGCGCTGGATACGCCCTCTTAC	1680
QY	561	PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys	580
DB	1681	TTTTCTTCTCTCTCTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1740
QY	581	AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys	600

Db 1741 GATGATAGTATCATCCAGTCCAGTGGACAACTGATCTGTAATCCGAGACCCCTCGTGTGT 1800  
 Qy 601 VallysGluProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAerGly 620  
 Db 1801 GTCAAGAGAGCCCGAGCTACACCTCCGCGACTACAGCGCCCATCTTTACCTAAATGGG 1860  
 Qy 621 AspPheAspGlyAenPheTyrPheThrGluLeuAspAlaIlysthrValThrAlaGlu 640  
 Db 1861 GACTTCGATCGCGAAACTTTTATTTTCACTGAATCGGACCGGAGCGGAGAG 1920  
 Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAenProHisGly 660  
 Db 1921 GTGCAGGCTCAGTGTGAAGAGCGGTGGGATTCTTTCAGGCACTGAAACCCACATGGA 1980  
 Qy 661 VallysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680  
 Db 1981 GTGAAGGCTGTACACAGGGGGCGCGCTGTGCCATCGCCCTGTGTGTTCACCTTGGACCT 2040  
 Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspLeuVallysMetLeuPheSerPro 700  
 Db 2041 CGACACAGCAGCGGAGCGGTGCGAGCAGATGACCTGTGGAAGTGTCTTCACGCCCA 2100  
 Qy 701 GluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
 Db 2101 GAAGAGATGACCTCTCCAGGAGCAGCCCTCGATGCCCGAGCAGGCTCCCCCGAACCT 2160  
 Qy 721 AlaGlnGluSerLeuSerGlySerGlySerGlySerGlySerGlySerGlySerGlySer 736  
 Db 2161 GCACAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGATGAGCTA 2208

RESULT 4  
 AAF93800  
 ID AAF93800 standard; cDNA; 2563 BP.

AC AAF93800;  
 XX  
 DT 23-MAY-2001 (first entry)  
 XX  
 DE Human cDNA encoding a membrane or secretory protein clone PSEC0109.  
 XX  
 KW Human; secretory protein; membrane protein; vaccine; gene therapy;  
 KW rheumatoid arthritis; diabetes; ss.  
 OS Homo sapiens.  
 XX  
 PN EPI067182-A2.  
 XX  
 PD 10-JAN-2001.  
 XX  
 PF 07-JUL-2000; 2000EP-00114090.  
 XX  
 PR 08-JUL-1999; 99JP-00194179.  
 PR 11-JAN-2000; 2000JP-00118775.  
 PR 02-MAY-2000; 2000JP-00183766.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;  
 XX  
 DR WPI; 2001-093989/11.  
 DR P-PSDS; AAB88373.  
 XX

XX Nucleic acids encoding secretory proteins/membrane proteins, useful in  
 PT gene therapy or as candidate target molecules in drug development.  
 XX  
 PS Claim 1; SEQ ID NO 113; 609pp + Sequence Listing; English.

XX This invention relates to nucleic acid sequences AAF93744 - AAF93916  
 CC which encode human secretory or membrane proteins represented by AAB88317  
 CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and  
 CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the  
 CC invention. The invention also includes methods for the production of

CC antibodies directed against the proteins, and cDNA sequences, which can  
 CC be used in vaccines. The polynucleotide sequences can be used in gene  
 CC therapy. The polynucleotide sequences and the proteins they encode may be  
 CC used in the prevention, treatment and diagnosis of diseases associated  
 CC with inappropriate secretory protein/membrane protein expression. The  
 CC nucleic acids and complementary sequences may also be used as DNA probes  
 CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect  
 CC and quantitate the presence of similar nucleic acid sequences in samples.  
 CC They may also be used to study the expression and function of secretory  
 CC proteins/membrane polypeptides and their role in metabolism. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC against them and in assays to identify modulators (agonists and  
 CC antagonists) of expression and activity. The antibodies and antagonists  
 CC may also be used as therapeutic agents to down regulate expression and  
 CC activity. The antibodies may also be used as diagnostic agents for  
 CC detecting the presence of the polypeptides in samples (e.g. by enzyme  
 CC linked immunosorbent assay (ELISA). Examples of diseases which may be  
 CC treated include rheumatoid arthritis and diabetes  
 XX

SQ Sequence 2563 BP; 589 A; 753 C; 713 G; 508 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 0 Length: 2563  
 Score: 3846.00 Matches: 733  
 Percent Similarity: 99.59% Conservative: 0  
 Best Local Similarity: 99.59% Mismatches: 3  
 Query Match: 99.38% Indels: 0  
 DB: 5 Gaps: 0

US-10-045-815-4 (1-736) x AAF93800 (1-2563)

Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValValAlaAlaSer 20  
 Db 31 ATGGCGGTACGCGGCTTGAAGCTGCTGACACACTGTGGCTGTGTGGCGCTGCCCTCC 90  
 Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuPheAla 40  
 Db 91 CAAGCCGAGGTGCGAGTCCGAGGAGGATGGGGGATGGAGCTGATCTGCTTCGCCC 150  
 Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60  
 Db 151 GAGGGGACCGCAGCCTACGCGCGCGGGACTGGCCCGGGGTGTCCTGAGCATGGAACGG 210  
 Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
 Db 211 GCGCTGCGCTCCCGGCGAGCCCTCCGCGCCCTTCGCTGCGCTCCGACCCAGTGTGCC 270  
 Qy 81 AlaAspPheProTyrGluLeuAspProAspTyrSerProSerProAlaGlnAlaSerGly 100  
 Db 271 GCCGACTTCCCGTGGGAGCTGACCCCGACTGTTCGCCAGCCGCGCCGCTCGGGC 330  
 Qy 101 AlaGlyAlaLeuA:GAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
 Db 331 GCGCGCGCCCTGCGCGACCTGAGCTTCTTCGGGGGCTTCTGCGCTGCGCTGCGCTG 390  
 Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe 140  
 Db 391 GCGCGCTGCTCGGGCGCGCGCCGCGCCACTGCTGCTGAGGAGAGATGGAGTGGAGTTC 450  
 Qy 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
 Db 451 CGCAAGCGGAGCCCTACCACTACCTGAGGTCCCTACTTCAAGATCAACAAGATTGGAG 510  
 Qy 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyValGlnProGluHisMetGluMetGln 180  
 Db 511 AAGCTGTGTGTCGAGCACACACCTTCTTCGTGGGCAATCTCTGAGCATGGAAATTCAG 570  
 Qy 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyVallysGluAlaAspPheLysAspLeu 200  
 Db 571 CAGAACCTTAGACTATTACCAAAACCATGTCTGGAGTGAAGGAGGCGGACTTCAAGGATCTT 630  
 Qy 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGln 220

631 GAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTGGCACTCTACTCAGAGGAACAG 690  
221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
691 CCACAGAGAGCTGTGGCCCACTAGAGGGGGCTGCAAGAAATCTTTGTGGCCCTATGAG 750  
241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
751 GAGTGGCGTGGCTCTGGAGGGCCCTATGACTACGATGGCTACAACTACCTTGATGATAC 810  
261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
811 AACGCTGACCTCTCCAGGCGCATCAGACATCATTACATCAGGCTCCTCAACTGTAAGCAG 870  
281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
871 AACTGTGTACGGAGCTGTCTCCCAAGTCGAGAGAGCCCTTTGAAGACTTCCTC 930  
301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
931 CCATCGCATTAATAATTATCTGCAGTTTGCCTACTATAACATTGGGAATTATACAGGCT 990  
321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
991 GTTGAATGTGCCAAGACCTATCTCTCTTCCCAATGAGAGGTGATGAACCAAAAT 1050  
341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluHisThrArgSerIleGlyProArgGlu 360  
1051 TTGGCTTATATGCAGCTATGCTTGGAGAGAACACACAGATCCATCGGCCCTCGTGAG 1110  
361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr 380  
1111 AGTGCCAAAGAGTACCGACAGCGAAGCTACTGCAAAAGAACTGCTTTCTCGCTTAT 1170  
381 AspValPheGlyLeuProPheValAspProAspSerTyrThrProGluGluValIlePro 400  
1171 GATGTTTTTGGAAATCCCTTTGGGATCCGGATTCATGGACTCCAGAAAGATGATTCCC 1230  
401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
1231 AAGAGATTGCAGAGAAACAGAGTCCAGAACGGAAACAGCGTACGCTATCCCGAGGAG 1290  
421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
1291 ATTGGGAACCTTATGAAGAAATCGAGAGCCCTTGTGGAAGAGAACCAAGAGAGTCACTG 1350  
441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
1351 GATGTGAGCAGACTGACCCGGAGAGTGGCCCTCTGTATGAAGGCATCAGTCTCACC 1410  
461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
1411 ATGAACCTCCAAACTCCTGAATGTTCCAGCGGGTGGTGTATGAGCGCGCTAATCTCTGAC 1470  
481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
1471 CAGGAGTGTGAGGAGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGCTTAC 1530  
501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
1531 CGGGGTACAGACTCCCAACATCTCCCAATGAAGTTCATGGTGTCACTGTCTTCAA 1590  
521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
1591 GCCTCAAGCTGGGGCAAGAACAAAGTTCCTCTCAGAGTGCCCACTGCTACTACAC 1650  
541 ValThrGluLysValArgGluMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
1651 GTACCGGAGAGTGGGGCGCATCTAGGAGTCTTACTTCCGCTGGATACGCCCTCTTAC 1710  
561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
1711 TTTTCTCTACTCTCATCTGTGTGCCGACTGCTCCATCGAAGAGTCCAGGAGAGGAG 1770

QY 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
Db 1771 GATGATAGTCAATCCAGTCCAGTCGACAACTGCATCTCGAATCCGAGACCTCGTGTGT 1830  
QY 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
Db 1831 GTCAAAGAGCCCCCAGCCTACACCTTCCGGACTACAGGGCCATCTTTACCTAAATGGG 1890  
QY 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
Db 1891 GACTTCGATGGCGAAACTTTTATTTTCACTGAATGGATGCCAAGACCGTGCAGCGCAGAG 1950  
QY 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
Db 1951 GTGAGACCTCACTGTGGAGAGCGTGGGATCTCTTCAGSCACTGAAACCCACATGGA 2010  
QY 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuAspPro 680  
Db 2011 GTGAAGGCTGTCAACAGGGGGCAGCGCTGTGCCATCGCCCTGTGTTTCACTCGACCT 2070  
QY 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
Db 2071 CGACACAGCGAGCGGAGCAGGGTGCAGCAGATGACCTGGTGAAGATGCTCTTCAGGCCA 2130  
QY 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
Db 2131 GAAGAGATGGACCTCTCCAGGAGCAGCCCTCGATGCCAGCAGGCGCCCCCGAACCT 2190  
QY 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2191 GCACAAGAGTCTCTCTCAGGCACTGAATCGAAGCCCAAGGATGAGCTA 2238

RESULT 5

AAH14589  
ID AAH14589 standard; cDNA; 2585 BP.  
XX  
AC AAH14589;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA sequence SEQ ID NO:12193.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-00116126.  
XX  
PR 29-JUL-1999; 99JP-00248036.  
PR 27-AUG-1999; 99JP-00300253.  
PR 11-JAN-2000; 2000JP-00118776.  
PR 02-MAY-2000; 2000JP-00183767.  
PR 09-JUN-2000; 2000JP-00241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Negai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
PS  
XX  
XX

Claim 8; SEQ ID NO 12193; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to a complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAB91632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 2585 BP; 590 A; 759 C; 723 G; 513 T; 0 U; 0 Other;

**Alignment Scores:**

red. No.:	0	Length:	2585
Score:	3846.00	Matches:	733
Percent Similarity:	99.59%	Conservative:	0
Best Local Similarity:	99.59%	Mismatches:	3
Every Match:	99.38%	Indels:	0
3:	4	Gaps:	0

3-10-045-815-4 (1-736) x AH14589 (1-2585)

1	MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaLeuSer	20
53	ATGCGGTGAGCGCTTGAAGCTGTGTGACCACTGCTGGCTGTGCTGGCGCTGCTGCTCC	112
21	GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla	40
113	CAAGCCGAGGTCCAGTCCGAGGACAGATGGGCATGGTGACCGCTGATCTGCTCTTCGCC	172
41	GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg	60
173	GAGGGGACCGACGCTACGCGCGCGGGACTGGCCCGGGTGGTCTTGACATGGGAACGG	232
61	AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla	80
233	CGCTGTGGCTCCCGGGACGCCCTCGCGGCTTGTGCTCGCTCCGCCACCCAGTGTGCC	292
81	AlaAspPheProTrpGluLeuAspProAspTrpSerProAlaGlnAlaSerGly	100
293	GCCGACTTCCGTGGAGAGCTGACCCCGACTGGTCCCCCAGCCGGCCGAGGGCTCGGGC	352
101	AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu	120
353	GCGCCGCCCTGCGCACCTGAGCTTCTTGGGGGCTTCTTCGTCGCGCTGCTGCTGCCGTG	412
121	ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluLeuMetGluLeuGluPhe	140
413	CGCCGTGCTCGGGCGCGCGGCCCCACCTGCTCAGCAAGAGATGGAGCTGGAGTTC	472
141	ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu	160
473	CGCAAGCGGAGCCCTACAACTACCTGCAGGTTCGCTTCTCAGATCAACAGTTGGAG	532
161	LysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetGln	180
533	AAAGCTGTTCCTCAGCACACACCTTCTTCGTGGGCAATCTCTGAGCATCTGGAAATGCG	592
181	GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu	200

```
QY 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArgLys 580
Db 1733 TTTTCTACTCTATCTGTTGTCGGCACTGCCATCGAGAGGTCAGGACAGAGAGAG 1792
QY 581 AspAspSerHisProValHisValAlaAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600
Db 1793 GATCATAGTTCATCCAGTCCACGTCGACAACTGCATCTGAATCCCGAGACCTCTCGTGTG 1852
QY 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620
Db 1853 GTCAAGAGCCCCAGGCTACACTTCCGGGATACAGGCCATCTTTTACCTAAATGGG 1912
QY 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640
Db 1913 GACTTCGATCGCGAAACTTTTATTTCACTGAATCGGATGCCAAGACCGTGACGGCAGAG 1972
QY 641 ValClnProClnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660
Db 1973 GTGCAGCCTCAGTGTGAAGAGCGGTGGGATTCTTTCAGGCACTGAAACCCACATGGA 2032
QY 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680
Db 2033 GTGAGGCTGTCAACAGGGGGGAGCGGTGTGCCATGCCCTGTGGTTACCTGGACCT 2092
QY 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700
Db 2093 CGACACAGCGAGCGGACAGCGGTGCAGGACAGTACCTGTGTGAAGATGCTTTCAGCCCA 2152
QY 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro 720
Db 2153 GAAAGATGAGCTTCTCCAGAGAGCAGCCCTGTGATGCCAGCGGGCCCCCGGACCT 2212
QY 721 AlaGlnGluSerLeuSerGlySerGlySerLysProLysAspGluLeu 736
Db 2213 GCACAAGAGTCTCTCTCAGCAGTGAATCGAAGCCCAAGATGAGCTA 2260

RESULT 6
AAH14513
ID AAH14513 standard; cDNA; 2524 BP.
XX AC
XX AAH14513;
XX DT
XX 26-JUN-2001 (first entry)
XX DE
XX Human cDNA sequence SEQ ID NO:12044.
XX KW
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX OS
XX Homo sapiens.
XX PN
XX EP1074617-A2.
XX XX
XX 07-FEB-2001.
XX XX
XX 28-JUL-2000; 2000EP-00116126.
XX PR
XX 29-JUL-1999; 99JP-00248036.
XX PR
XX 27-AUG-1999; 99JP-00300253.
XX PR
XX 11-JAN-2000; 2000JP-00118776.
XX PR
XX 02-MAY-2000; 2000JP-00183767.
XX PR
XX 09-JUN-2000; 2000JP-00241899.
XX XX
XX (HELI-) HELIX RES INST.
XX PA
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX DR
XX WPI; 2001-318749/34.
XX XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
XX PT
XX length cDNAs defined in the specification, and for the detection and/or
XX PT
XX diagnosis of the abnormality of the proteins encoded by the full-length
```

```
PT cDNAs.
XX Claim 8; SEQ ID NO 12044; 2537pp + Sequence Listing; English.
XX CC
XX The present invention describes primer sets for synthesizing 5602 full-
XX CC
XX length cDNAs defined in the specification. Where a primer set comprises:
XX CC
XX (a) an oligo-dT primer and an oligonucleotide complementary to the
XX CC
XX complementary strand of a polynucleotide which comprises one of the 5602
XX CC
XX nucleotide sequences defined in the specification, where the
XX CC
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX CC
XX of an oligonucleotide comprising a sequence complementary to the
XX CC
XX complementary strand of a polynucleotide which comprises a 5'-end
XX CC
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX CC
XX polynucleotide which comprises a 3'-end sequence, where the
XX CC
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX CC
XX the 5'-end sequence/3'-end sequence is selected from those defined in the
XX CC
XX specification. The primer sets can be used in antisense therapy and in
XX CC
XX gene therapy. The primers are useful for synthesizing polynucleotides,
XX CC
XX particularly full-length cDNAs. The primers are also useful for the
XX CC
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX CC
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX CC
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX CC
XX AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
XX CC
XX represent human amino acid sequences; and AAH13629 to AAH13632 represent
XX CC
XX oligonucleotides, all of which are used in the exemplification of the
XX CC
XX present invention
XX SQ
XX Sequence 2524 BP; 583 A; 747 C; 704 G; 490 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.: 0 Length: 2524
Score: 3844.00 Matches: 732
Percent Similarity: 99.5% Conservative: 1
Best Local Similarity: 99.46% Mismatches: 3
Query Match: 99.33% Indels: 0
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x AAH14513 (1-2524)
```

```
QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaAlaSer 20
Db 16 ATGGCGGTACGGCGCTTGAAGCTGCTGACCACTGCTGGGTGTGTCGGCGCTGCCCTCC 75
QY 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla 40
Db 76 CAAAGCCGAGGTGAGTCCGAGGAGATGGGCGATGGTGACGCTGATCTGCTCTCCGCC 135
QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60
Db 136 GAGGGGACCGCAGCCTACGCGCGGGGACTGGCCCGGGTGTCTCTGAGCATGGAACGG 195
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
Db 196 GCGCTGCGCTCCGGGCGAGCCCTCCGCGCCCTTCCCTGCGCTGCCGACCCAGTGTGCC 255
QY 81 AlaAspPheProTyrGluLeuAspProAspTyrProSerProAlaGlnAlaSerGly 100
Db 256 GCCGACTTCCCGTGGAGCTGGACCCGACTGGTCCCGCAGCCCGCCAGGCTCGGGC 315
QY 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLeu 120
Db 316 GCCGCGCGCTCCGGGCGAGCTTCTTCGGGGGCCCTTTCGCGCGGCTCCCTGCGCTG 375
QY 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe 140
Db 376 CGCCGCTGCTCGCGCGCCCGCCCGCCCTCGCTCAGGAGAGATGAGCTGAGTTC 435
QY 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160
Db 436 CGCAAGCGGAGCGCCCTACAACTACCTGAGGTGCGCTACTTCAAGATCAACAGTTGAG 495
QY 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180
```

496 AAAGCTGTGTCGACGACACACCTTCTCTGTCGGCAATCTCTGAGCACATGGAAATGCAG 555  
181 GlnAsnLeuAspTyrGlnThrMetSerGlyValIysGluAlaAspPheLysAspLeu 200  
556 CAGAACTAGACTATTACCAAAACATGTCGAGTGAAGGAGGCGGACTTCAAGGATCTT 615  
201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220  
616 GAGACTCAACCCCATATCAAGAATTTGACTGGAGTGGGACTCTACTCAGAGGAACAG 675  
221 ProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
676 CCACAGGAAGCTGTGCCCCCCTAGAGGCGCGCTGCAAGAATACTTTGTGGCCTATGAG 735  
241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
736 GAGTCCCGTGGCCCTCTGCAAGGCGCCCTATGACTACGATGCTACAACTACTTGAATAC 795  
261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
796 AACGCTGACCTTCTCCAGGCCCATCACAGATCATTAATCCAGGTCTCAACTGTAAACAG 855  
281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
856 AACTGTGTACGGAGCTGTCTTCCACCCCAAGTCGAGAGAGCCCTTTGAAGACTTTCCTC 915  
301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
916 CCATCGCATTAATAATATCTCGAGTTGCCCTACTATAACATTGGGAATTATACACAGGCT 975  
321 GlyGluCysAlaIysThrTyrLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
976 GTTGAATGTGCAAGACCTATCTCTCTTCCCAATGACGAGGTGATGAACCAAAAT 1035  
341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluHisThrArgSerIleGlyProArgGlu 360  
1036 TTGGCCCTATTATGAGCTATGCTTGGAGAGAAACACACACAGATCCATCGGCCCCCGTGAG 1095  
361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhePheAlaTyr 380  
1096 AGTCCCAAGAGTACCCAGACGAGCTACTGGAAGAAAGAACTGCTTTTCTTCGCTTAT 1155  
381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400  
1156 GATGTTTTTGGAAATTCCTTTGTGGATCCGAGATTATGAGTCCAGAAAGATGATTCCTC 1215  
401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
1216 AAGAGATTGCAAGAGAAACAGAGTCAAGACGGGAAACACCGGTACGCATCTCCAGGAG 1275  
421 IleGlyAsnLeuMethLysGluLeuThrLeuValGluGluLysThrLysGluSerLeu 440  
1276 ATTGGGAACCTTATGAAGAAATCGACACCTTGTGGAAGAGAACCAAGGAGTCACTG 1335  
441 AspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
1336 GATGTGAGCAGACTGACCGGAGAGTGGCCCTGCTGTATGAAGGATCATGCTCAACC 1395  
461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
1396 ATGAACCTCAAACTCTGTAATGTTGTCACGCGGTGTGTGACGCGCTAATCTCTGAC 1455  
481 HisGlnCysGlnGluLeuGlnArgLeuThrAsnValAlaIleThrSerGlyAspGlyTyr 500  
1456 CACAGAGTGTGAGGAGTGCAGAGACTGACCAATGTGCGACAACTCCAGAGATGCTCTAC 1515  
501 ArgGlyGlnThrSerProHisThrProAsnGlnLysPheTyrGlyValThrValPheLys 520  
1516 CGGGTTCAGACCTCCCCACATCTCCCAATGANAAGTTCTATGTCATGCTGCTTCAAA 1575  
521 AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
1576 GCCTCAAGCTGGGGCAAGAGCAAGTTCCTCTGACAGTCCCAACCTGTACTACAAC 1635

541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
1636 GTGACGGAGAGGTGGCGGCATCATGAGTCTCTACTTCCGCTGGATACGCCCTCTAC 1695  
561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGlnArgLys 580  
1696 TTTTCT 1755  
581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
1756 GATGATAGTCAATCCAGTCCACGTGCACAACTGCATCTCTGAATGCCGACACCTGTGTGT 1815  
601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
1816 GTCAAAGAGCCCCCAGCCTACACCTTCCGCGACTACAGCGCATCTTCTTAAATGGG 1875  
621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
1876 GACTTCGATGGCGAACTTTTATTTCTACTGACTGATGCCAGACCTGTACGCGCAGAG 1935  
641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
1936 GTGCAGCCTCAGTGTGGAAGAGCCGTGGGATTCTCTCAGGCACCTGAAACCCACATGGA 1995  
661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680  
1996 GTGAAGCTGTCCACAGGGGCGAGCGCTGTGCCATCGCCCTGTGTTCCACCTTGACCCCT 2055  
681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
2056 CGACACAGCAGCGGACAGGCTGCGAGCAGATGACCTGTGTGAAGATGCTCTTCAAGCCA 2115  
701 GluGluMetAspLeuSerGlnGluProLeuAspAlaGlnGlnGlyProProGluPro 720  
2116 GAAGAGATGGACCTCTCCAGAGCAGCCATGATGCCAGAGGCGGCCCCCGAAACCT 2175  
721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
2176 GCACAGAGTCTCTCTCAGCAGTGAATCGAAGCCCAAGGATGAGCTA 2223

## RESULT 7

AAD08489

ID AAD08489 standard; cDNA; 2577 BP.

XX AC AAD08489;

XX XX 09-AUG-2001 (first entry)

XX XX Human secreted protein-encoding gene 2 cDNA clone HUVFY29, SEQ ID NO:12.

XX XX Human; secreted protein; proliferative disorder; cancer; tumour;

XX XX foetal abnormality; developmental abnormality; haematopoietic disorder;

XX XX immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;

XX XX inflammation; allergy; neurological disorder; Alzheimer's disease;

XX XX Parkinson's disease; cognitive disorder; schizophrenia; asthma;

XX XX skin disorder; psoriasis; sepsis; diabetes; kidney disease;

XX XX cardiovascular disorder; angiotensin-related disorder; gene therapy;

XX XX endocrine disorder; infection; wound healing; vulnerability; cell culture;

XX XX chemotaxis; food additive; binding partner identification; ss.

XX XX Homo sapiens.

XX XX Key Location/Qualifiers

XX XX 24..1103 /tag= a

XX XX /product= "Human secreted protein precursor"

XX XX 24..77 /tag= b

XX XX 78..1100 /tag= c

XX XX /product= "Mature human secreted protein"

XX XX sig\_peptide

XX XX mat\_peptide

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX	PN	WO200136432-A2.	DB	144	GAGGGGACCGAGCCTACGGCGCGGGGACTGGCCCGGGGTGGTCTCTGAGCATGGAACGG	203
XX	PD	25-MAY-2001.	QY	61	AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgCysArgThrGlnCysAla	80
XX	PF	15-NOV-2000; 2000WO-US031162.	DB	204	GGCTGGCTCCGGGCGAGCCCTCCGGCCCTTCGCTCGCTGCCGCGCAGTGTGCC	263
XX	PR	19-NOV-1999; 99US-0166415P.	QY	81	AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly	100
XX	PR	30-JUN-2000; 2000US-0215136P.	DB	264	GCCGACTTCCGCTGGGAGCTGGACCCGACTGGTCCCGCAGCCGCGCCAGGCTCGGGC	323
XX	PA	(HUMA-) HUMAN GENOME SCI INC.	QY	101	AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLeu	120
XX	PI	Ruben SM, Komatsoulis GA, Baker KP, Young PE;	DB	324	GCGCGCGCTTGGCGGACCTGAGCTTCTTCGGGGGCGCTTCGCTGCGCTGCCTG	383
XX	WPI; 2001-343793/36.	QY	121	ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe	140	
XX	P-PSDB; AAE04200.	DB	384	CGCCGCTGCTCGGCGCGCGCCGCGCCACTCGCTCAGCGAAGAGATGAGCTGGAGTTC	443	
PT	Isolated nucleic acid molecule encoding a human secreted protein is used	QY	141	ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu	160	
PT	in preventing, treating or ameliorating a medical condition.	DB	444	CCCAAGCGGAGCCCTCACTACTCTGCGAGTGGCTACTTCAAGATCAACAAGTTGGAG	503	
PS	Claim 1; Page 403-404; 509pp; English.	QY	161	LysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetGln	180	
CC	AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted	DB	504	AAAGCTGTGTGTCGACGACACACACTTCTTCGTGGGCATCTCTGAGCACATGGAATGCG	563	
CC	protein genes, and AAE04199-AAE04239 represent the proteins they encode.	QY	181	GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu	200	
CC	AAE04240-AAE04277 represent human secreted protein fragments or variants.	DB	564	CAGAACCTAGACTATTACCAACCATGTCTGAGTGAAGGAGGCGCGACTTCAAGGATCTT	623	
CC	The secreted proteins and their genes are useful for preventing, treating	QY	201	GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln	220	
CC	or ameliorating medical conditions, e.g., by protein or gene therapy.	DB	624	GAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTGGGACTCTACTCAGAGGAACAG	683	
CC	Pathological conditions can be diagnosed by determining the amount of the	QY	221	ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu	240	
CC	new protein in a sample or by determining the presence of mutations in	DB	684	CCACAGGAGCTGTGCCCCACCTAGAGCGCGCGCTGCAAGAATACTTTGGCGCTATGAG	743	
CC	the 18 genes. Specific uses are described for each of the 18 genes,	QY	241	GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr	260	
CC	based on the tissues in which they are most highly expressed, and include	DB	744	GAGTCCGCTGCGCTCTGCAAGGCGCCCTATGACTACGATGGCTACAACTTCTGAGTAC	803	
CC	developing products for the diagnosis or treatment of proliferative	QY	261	AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln	280	
CC	disorders, cancer, tumours, foetal and developmental abnormalities,	DB	804	AACGCTGACCTTCTCCAGGCCATCAGAGATCATATCATCCAGTCTCTCAACTGTAAAGCAG	863	
CC	haematopoietic disorders, diseases of the immune system, AIDS, autoimmune	QY	281	AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu	300	
CC	diseases (e.g., rheumatoid arthritis), inflammation, allergies,	DB	864	AACTGTGTACGAGCTTGTCTCCACCCAAAGTCGAGAGAGGCCCTTTGAGACTTCTCTC	923	
CC	neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),	QY	301	ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla	320	
CC	cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,	DB	924	CCATCGCATTAATATCTGCAGTTTTCGCTACTATAACATTTGGGAATTTATACACAGGCT	983	
CC	psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,	QY	321	GlyGluCysAlaAlaValThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn	340	
CC	angiogenic disorders, kidney disorders, gastrointestinal disorders,	DB	984	GTTGAATGTGCCAAGACCTATCTTCTTCTTCCCAATGACGAGTGTATGAACCAAAAT	1043	
CC	pregnancy-related disorders, endocrine disorders, and infections. The	QY	341	LeuAla-TyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGln	360	
CC	proteins can also be used to aid wound healing and epithelial cell	DB	1044	TTGGCCCTATTATGCGACTATGCTTTGGAGAGAAACACACAGATCCATCGGCCCGCTGA	1103	
CC	proliferation, to prevent skin ageing due to sunburn, to maintain organs	QY	360	uSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr	380	
CC	before transplantation, for supporting cell culture of primary tissues,	DB	1104	GAGTGCACAGGAGTACCGACAGCAAGCTACTTGGAAAAGAACTGCTTTTCTTCGCTTA	1163	
CC	to regenerate tissues, to identify their cognate ligands or binding	QY	380	rAspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePr	400	
CC	partners, and in chemotaxis, and can be used as a food additive or	DB	1164	TGATGTTTTTGGAAATTCCTTTGTGGATCCGGAATTCATGGACTCCAGAGAGTGAITTC	1223	
CC	preservative to modify storage properties. Antibodies specific for a	QY	400	olysArgLeuGlnGlu-LysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG	420	
CC	protein of the invention can be used in alleviating symptoms associated					
CC	with the disorders mentioned above, and in diagnostic immunoassays e.g.,					
CC	radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The					
CC	present sequence represents a human secreted protein-encoding cDNA of the					
CC	invention					
SQ	Sequence 2577 BP; 610 A; 753 C; 711 G; 503 T; 0 U; 0 Other;					
Alignment Scores:						
Pred. No.:	0	Length:	2577			
Score:	3820.00	Matches:	733			
Percent Similarity:	99.32%	Conservative:	0			
Best Local Similarity:	99.32%	Mismatches:	3			
Query Match:	98.71%	Indels:	2			
DB:	4	Gaps:	0			
US-10-045-815-4 (1-736) x AAD08489 (1-2577)						
QY	1	MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer	20			
DB	24	ATGGCGGTACCGCGTGTGAAGCTGCTGACCACACTGCTGGCTGTCGTGGCGCGCTGCCCTC	83			
QY	21	GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla	40			
DB	84	CAACCCGAGGTGAGTCCGAGCGAGGATGGGCATGTTGACCGCTGATCTGCTCTTCGCC	143			
QY	41	GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg	60			



Db 1224 CAGAGATTGCAAGAGAAACAGAAAGTCAGAACCGGAAACAGCGGTACGCACTCTCCAGG 1283  
 QY 420 luileGlyAsnLeuMetLysGluThrLeuValGluGluLysThrLysGluSerL 440  
 Db 1284 AGATTGGGAACCTTATGAAGAAATCGAGACCCCTGTGTGAAGAGAGACCAAGGAGTCAC 1343  
 QY 440 euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuThrGluGlyLysLeuLeu 460  
 Db 1344 TGGATGTGAGCAGACTGACCCGGAAGGTGGCCCTGCTGTATGAAGGCACTAGTCTCA 1403  
 QY 460 hrMetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValLysSerA 480  
 Db 1404 CCATGAATCCCAACTCTGAATGTTCCCGAGCGGTGGTGTGATGAGCGGTATCTCTG 1463  
 QY 480 spHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyT 500  
 Db 1464 ACCACGAGTGTGAGGAGTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGCT 1523  
 QY 500 yrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheL 520  
 Db 1524 ACCGGGTTCAGACTCCCACTACTCCCAATGAAGATTCATGGTGTCACTGTCTTCA 1583  
 QY 520 ysAlaLeuLysGluGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrA 540  
 Db 1584 AAGCCCTCAAGCTGGGCAAGAGCAAAAGTTCCTCTGCAGAGTGGCCCACTGTACTACA 1643  
 QY 540 snValThrGluLysValArgArgileMetGluSerTyrPheArgLeuAspThrProLeuT 560  
 Db 1644 ACGTGACGGAAGGTGGCGGATCATGGAGTCTTCTCCGCTGGATACGCCCTCT 1703  
 QY 560 yrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgL 580  
 Db 1704 ACTTTTCTACTCTCATCTGTGTGGCGCACTGCCATCGAAGAGTCCAGCGAGAGGA 1763  
 QY 580 ysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValC 600  
 Db 1764 AGGATGATAGTCATCCAGTCACGTGACAACTGCATCTCCGTAATGCGGACCCCTGTGT 1823  
 QY 600 ysValLysGluProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnG 620  
 Db 1824 GTGTCAAGAGCCCCAGCTACACCTTCGCGACTACAGCGCATCTTTACCTAAATG 1883  
 QY 620 lyAspPheAspGlyClyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaG 640  
 Db 1884 GGGACTTCGATGGCGGAACCTTTTATTTCACTGAACCTGGATGCCAAGACCGTGACGCG 1943  
 QY 640 luValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisG 660  
 Db 1944 AGTGTGACCTCAGTGTGGAAGACCGTGGATCTCTTTCAGGCATGAAACCCACATG 2003  
 QY 660 lyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspP 680  
 Db 2004 GAGTCAAGGCTGTCCACGAGGGGCGGCTGTGCACTGGCCCTGTGTTTCACTCGACC 2063  
 QY 680 roArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerP 700  
 Db 2064 CTCGACACAGGCGGAGAGGTGGAGGATGACCTGGTGAAGATGCTCTTCAGCC 2123  
 QY 700 roGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluP 720  
 Db 2124 CAGAAGAGATGACCTCTCCAGGAGCAGCCCTGGATGCCAGCGGGCCCCCCCCGAC 2183  
 QY 720 roAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
 Db 2184 CTGCACAGAGTCTCTCTCAGCGAGTGAATCGAAGCCCAAGGATGAGCTA 2233

RESULT 8

ABK94910

ID ABK94910 standard; cDNA; 2753 BP.

XX ABK94910;

AC ABK94910;

DT 30-AUG-2002 (first entry)

XX DE Human novel polynucleotide #21.  
 XX KW Human; gene; ss; inflammatory condition; shock; sepsis; immune response;  
 KW cancer; wound healing; central nervous system disease; haematopoiesis;  
 KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;  
 KW myeloid cell disorder; lymphoid cell disorder; placental disorder; bone;  
 KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;  
 KW bone degenerative disorder; periodontal disease; reperfusion injury;  
 KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;  
 KW allergic condition; thrombolysis; thrombosis; coagulation disorder;  
 KW fungal infection.  
 XX KW Homo sapiens.  
 OS WO200244340-A2.  
 XX PN 06-JUN-2002.  
 XX PD 30-NOV-2001; 2001WO-US047004.  
 XX PF 30-NOV-2000; 2000US-00728952.  
 XX PR (HYSE-) HYSEQ INC.  
 XX PA Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;  
 XX PI Yamazaki V, Ujwal ML, Drmanac RT;  
 XX PT WPI; 2002-508509/54.  
 XX PS P-PSDB; ABG66686.  
 XX DR Novel nucleic acids and polypeptides for diagnosis, treatment of  
 XX PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell  
 XX PT disorders, cancer and promoting wound healing.  
 XX PS Claim 1; Page 369-373; 672pp; English.

The invention relates to human novel polynucleotides and associated  
 polypeptides. The polynucleotides and polypeptides are useful for  
 treating inflammatory conditions such as arthritis, nephritis, Crohn's  
 disease, ischemia-reperfusion injury, shock, sepsis, immune responses  
 and cancer and for promoting wound healing. The sequences are used to  
 induce the proliferation of neural cells and regeneration of nerve and  
 brain tissue, and are useful for the treatment of central and peripheral  
 nervous system diseases and neuropathies, such as Alzheimer's disease,  
 Parkinson's disease, Huntington's disease and amyotrophic lateral  
 sclerosis. The sequences are involved in chemotactic or chemokinetic  
 activity, regulation of haematopoiesis, treatment of myeloid or lymphoid  
 cell disorders and platelet disorders such as thrombocytopenia,  
 regeneration of bone, cartilage, tendon, ligament and/or nerve tissue  
 growth, tissue repair, healing of burns, incisions, ulcers, treatment of  
 osteoporosis, osteoarthritis, bone degenerative disorders and periodontal  
 disease. The sequences of the invention are also useful for gut  
 protection or regeneration and treatment of lung or liver fibrosis.  
 CC reperfusion injury in various tissues, immune deficiencies and disorders  
 CC including severe combined immunodeficiency (SCID), bacterial or fungal  
 CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia  
 CC gravis, allergic conditions such as asthma, thrombolysis or thrombosis  
 CC and coagulation disorders. Sequences ABK94890-ABK94982 represent human  
 CC novel polynucleotides of the invention

SQ Sequence 2753 BP; 632 A; 802 C; 784 G; 535 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0	Length:	2753
Score:	3801.00	Matches:	733
Percent Similarity:	90.94%	Conservative:	0
Best Local Similarity:	90.94%	Mismatches:	3
Query Match:	98.22%	Indels:	70
DB:	6	Gaps:	1

US-10-045-815-4 (1-736) x ABK94910 (1-2753)





Db 2161 TTCTCTCAGGACCTGAAACCCACCATGAGTGAAGCTGTACACCGGGGCGAGCGTGT 2220  
 Qy 671 AlalleAlaLeuTTPreThrLeuAspProArgHisSerGluArgAspArgValGlnAla 690  
 Db 2221 GCCATCGCCCTGTGTTCACCTCGACCTCGACACGAGCGGACAGGGTGCAGGCA 2280  
 Qy 691 AspAspLeuValLysMetLeuPheSerProGluGluMetAspLeuSerGlnGluGlnPro 710  
 Db 2281 GATGACCTGGTGAAGATGCTCTTCAGCCACAGAGATGACCTCTCCAGGAGCAGCCC 2340  
 Qy 711 LeuAspAlaGlnGlnGlyProProGluProAlaGlnGluSerLeuSerGlySerGluSer 730  
 Db 2341 CTGGATGCCCGACGAGCGGCCCCCGAACCTGACACAGAGTCTCTCTCAGGACGTGAATCG 2400  
 Qy 731 LysProLysAspGluLeu 736  
 Db 2401 AAGCCCAAGGATGAGCTA 2418

RESULT 9

ID ABK94933  
 XX ABK94933 standard; cDNA; 2753 BP.

AC ABK94933;

XX 30-AUG-2002 (first entry)

DE Human novel polynucleotide #44.

XX Human; gene; ss; inflammatory condition; shock; sepsis; immune response;  
 KW cancer; wound healing; central nervous system disease; haematopoiesis;  
 KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;  
 KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;  
 KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;  
 KW bone degenerative disorder; periodontal disease; reperfusion injury;  
 KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;  
 KW allergic condition; thrombolytic; thrombosis; coagulation disorder;  
 KW fungal infection.

XX Homo sapiens.

XX WO200244340-A2.

XX 06-JUN-2002.

XX 30-NOV-2001; 2001WO-US047004.

XX 30-NOV-2000; 2000US-00728952.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;  
 PI Yamazaki V, Ujwal ML, Drmanac RT;

XX WPI; 2002-508509/54.

XX P-PSDB; ABG66709.

PT Novel nucleic acids and polypeptides for diagnosis, treatment of  
 PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell  
 PT disorders, cancer and promoting wound healing.

XX Claim 1; Page 441-444; 672pp; English.

XX The invention relates to human novel polynucleotides and associated  
 CC polypeptides. The polynucleotides and polypeptides are useful for  
 CC treating inflammatory conditions such as arthritis, nephritis, Crohn's  
 CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses  
 CC and cancer and for promoting wound healing. The sequences are used to  
 CC induce the proliferation of neural cells and regeneration of nerve and  
 CC brain tissue, and are useful for the treatment of central and peripheral  
 CC nervous system diseases and neuropathies, such as Alzheimer's disease,  
 CC Parkinson's disease, Huntington's disease and amyotrophic lateral  
 CC sclerosis. The sequences are involved in chemotactic or chemokinetic  
 CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid

CC cell disorders and platelet disorders such as thrombocytopenia,  
 CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue  
 CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of  
 CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal  
 CC disease. The sequences of the invention are also useful for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues, immune deficiencies and disorders  
 CC including severe combined immunodeficiency (SCID), bacterial or fungal  
 CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia  
 CC gravis, allergic conditions such as asthma, thrombolytic or thrombotic  
 CC and coagulation disorders. Sequences ABK94890-ABK94982 represent human  
 CC novel polynucleotides of the invention  
 XX  
 SQ Sequence 2753 BP; 632 A; 802 C; 784 G; 535 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 2753  
 Score: 3801.00 Matches: 733  
 Percent Similarity: 90.94% Conservative: 0  
 Best Local Similarity: 90.94% Mismatches: 3  
 Query Match: 98.22% Indels: 70  
 DB: 6 Gaps: 1

US-10-045-815-4 (1-736) x ABK94933 (1-2753)

Qy 1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaSer 20  
 Db 1 ATGGCGGTACGGCGCTTGAAGCTGTGACACACACTGCTGGCTGTGCGCGCTGCCTCC 60  
 Qy 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40  
 Db 61 CAAGCCGAGGTGAGTCCGAGCGAGATGGGCGATGGAGCGCTGATCTCTTTCGCC 120  
 Qy 41 GluGlyThrAlaAlaTyAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
 Db 121 GAGGGGACCGCAGCTACGCGCGGGGACTGGCCCGGGTGGTCTTGAGCATGGACGG 180  
 Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
 Db 181 GCGCTGCGCTCCCGGGCAGCCCTCGCGCCCTTGGCTGCGCTCGCGCACCGATGTGCC 240  
 Qy 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
 Db 241 GCGGACTTCCCGTGGAGCTGGACCCGACTGGTCCCCCAGCCGCCCGCCCTCGGGC 300  
 Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLeu 120  
 Db 301 GCGCGCGCCCTCGCGACCTGAGCTTCTTCGGGGGCGCTTCTGCGTGGCGCTGCGCTG 360  
 Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
 Db 361 CCGCCCTGCTTGGCGCGCCCGCCACCTGCTCAGCGAAGAGATGGAGCTGGAGTTC 420  
 Qy 141 ArgLysArgSerProTyAsnTyLeuGlnValAlaTyPhe----- 154  
 Db 421 CGCAAGCGGAGCCCTACAACTACTGCGAGTGCCTACTTCAAGGTGCAGACCTGCTG 480  
 Qy 154 ----- 154  
 Db 481 GAACAGCGCGCGGGGTCTTTCGGGGAGAGAGTGTTCAGGGGACCTGAGGAGTTG 540  
 Qy 154 ----- 154  
 Db 541 GGGGATCGGGAGAGTGTCCGCGAGGGAGGAAAGTGGCTCTCTGCTGGGAGCTCTCT 600  
 Qy 154 ----- 154  
 Db 601 CGGAGCCGGGAGAGTGTCTCCCTGGCAGGAGACCTTCTCTCGCCAGTTCGATGGGAG 660  
 Qy 155 -----LysIleAsnLysLeuGluLysAlaValAlaAlaHisThrPhePhe 170  
 Db 661 ATGCTAACCCCAAGATCAACAAGTTGGAGAAAGCTGTGTGTCAGCACACACCTTCTTC 720

Qy	171	ValGlyAsnProGluHisMetGluMetGlnGlnAsnLeuAspTyrTyrGlnThrMetSer	190
Db	721	GTGGGCAATCCTGAGCACATGGAAATGCAGCAGAACCTAGACTATTACCAACCATGTCT	780
Qy	191	GlyValLysGluAlaAspPheLysAspLeuGluThrGlnProHisMetGlnGluPheArg	210
Db	781	GGAGTGAAGAGGCGGCACTTCAAGGATCTTTGAGACTCACCCCATATGCAGAAATTTGGA	840
Qy	211	LeuGlyValArgLeuTyrSerGluGluGlnProGlnGluAlaValProHisLeuGluAla	230
Db	841	CTGGAGTGCAGCTCTACTCAGAGGAACAGCCACAGGAAGCTGTGCCCCACCTAGAGGCG	900
Qy	231	AlaLeuGlnGluTyrPheValAlaTyrGluGluCysArgAlaLeuCysGluGlyProTyr	250
Db	901	CGCGTCGAAGAAATCTTTGTGGCCTATGAGGAGTGGCGTCCCTCTGCGAAGGGCCCTAT	960
Qy	251	AspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAlaIleThrAsp	270
Db	961	GACTACGATGGCTACAACTACCTTGAGTCAACGCTGACCTCTTCCAGGCCATCAGAT	1020
Qy	271	HisTyrIleGlnValLeuAsnCysLysGlnAsnCysValThrGluLeuAlaSerHisPro	290
Db	1021	CATTACATCCAGGTCTCTCAACTGTAACAGCAACTGTGTACGAGAGCTGTCTTCCCA	1080
Qy	291	SerArgGluLysProPheGluAspPheLeuProSerHisTyrAsnTyrLeuGlnPheAla	310
Db	1081	AGTCGAGAGAGCCCTTTGAAGACTTCTCCCATCGCAATTATATATATCTGCAAGTTGCC	1140
Qy	311	TyrTyrAsnIleGlyAsnTyrThrGlnAlaGlyGluCysAlaLysThrTyrLeuLeuPhe	330
Db	1141	TACTATTAACATTGGGAATTATACACAGGCTGTTGAATGTGCCAAGACCTATCTTCTCTTC	1200
Qy	331	PheProAsnAspGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGlu	350
Db	1201	TTCCCCAATGACGAGGTGATGAACCAAAATTTGGCCATTATTGCAGCTATGCTTGAGAA	1260
Qy	351	GluHisThrArgSerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeu	370
Db	1261	GAACACACCAAGATCCATCGGCCCCCGTGAGAGTGCCCAAGCAGTACGACAGCGAA	1320
Qy	371	LeuGluLysGluLeuLeuPhePheAlaTyrAspValPheGlyIleProPheValAspPro	390
Db	1321	CTGAAAGAAAGCAATGCTTTTCTTCGCTTATGATGTTTTTGGAAATTCCTTTGTGGATCCG	1380
Qy	391	AspSerTrpThrProGluGluValIleProLysArgLeuGlnGluLysGlnLysSerGlu	410
Db	1381	GATTCAATGGATCCAGAAAGAAAGTGATTCCCAAGAGATTGCAAGAGAAACAGAA	1440
Qy	411	ArgGluThrAlaValArgIleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThr	430
Db	1441	CGGGAAACAGCGTACGCATCTCCAGGAGATTGGGAACCTTATGAAGGAAATTCGAGACC	1500
Qy	431	LeuValGluLysThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyGly	450
Db	1501	CTTGTGAGAGAGAACCAAGGAGTCACTGGATGTGAGCAGACTGACCCCGGAAGGTGGC	1560
Qy	451	ProLeuLeuTyrGluGlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyrGln	470
Db	1561	CCCCGTGCTGATGAAGGCATCAGTCTCACCATGAACCTCCAAACTCTCTGAATGTCTCCAG	1620
Qy	471	ArgValValMetAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThr	490
Db	1621	CGGTTGTGATGAGCGCGTAATCTCTGACCAACAGAGTGTGAGGAGCTGACAGACTGACC	1680
Qy	491	AsnValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsn	510
Db	1681	AATGTGGCAGCAACCTCAGGAGATGGGTATCGGGGGTCAGACCTCCCCACATATCTCCCAAT	1740
Qy	511	GluLysPheTyrGlyValThrValPheIleAlaLeuLysLeuGlyGlnGluGlyLysVal	530
Db	1741	GAAAGATTCTATGTGTGTCATGTCCTTCAAGCCCTCAAGCTGGGGGCAAGAGCAAGTT	1800
Qy	531	ProLeuGlnSerAlaHisLeuTyrTyrAsnValThrGluLysValArgArgLeuMetGlu	550

		CCTCTGCAGAGTCCCACTGTACTAACGTCACGGAGAAGTGCGGCGCATCGGAG	1860
Db			
Qy	551	SerTyrPheArgLeuAspThrProLeuTyrPheSerTyr-SerHisLeuValCysArgThr	570
Db	1861	TCTACTTCGCGCTGGATAGCCCCCTCTACTTTTCTCTACTCTCATCTGGTGTGCCGCACT	1920
Qy	571	AlaIleGluValGlnAlaGluArgLysAspSerSerHisProValHisValAspAsn	590
Db	1921	GCCATCGAAGAGGTCCAGGCAGAGAGGAAGATGATGTCATCCAGTCCAAGTGGACAAC	1980
Qy	591	CysIleLeuAsnAlaGluThrLeuValCysValLysGluProProAlaTyrThrPheArg	610
Db	1981	TGCATCTGTANTGCCGNGACCCTCGTGTGTCTCAAGAGCCCCAGCCTACACTTCCGC	2040
Qy	611	AspTyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThr	630
Db	2041	GACTACAGCCCATCTCTTACCTAAATGGGGAATTGCGATGCGGAAAACTTTTATTTCAC	2100
Qy	631	GluLeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGly	650
Db	2101	GAATCTGATGCCAAGACCCTGACGGCAGAGGTGCGACCTCAGTGTGGAGAGCGGTGGGA	2160
Qy	651	PheSerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCys	670
Db	2161	TTCTCTTCAGGCACTGAAAAACCCACATGGAGTGAAGGCTGTCCACAGGGGGCAGCGTGT	2220
Qy	671	AlaIleAlaLeuTrpPheThrIleuAspProArgHisSerGluArgAspArgValGlnAla	690
Db	2221	GCCATCGCCCTGTGTTCACCTCGGACCCCTCGACAGCGAGCGGCACAGGGTGCAGGCA	2280
Qy	691	AspAspLeuValLysMetLeuPheSerProGluGluMetAspLeuSerGlnGluInPro	710
Db	2281	GATGACCTGTGAAGATGCTCTTCAGCCCAGAGAGATGGACCTCTCTCCAGGAGCAGCCC	2340
Qy	711	LeuAspAlaGlnGlnGlyProProGluProAlaGlnGluSerLeuSerGlySerGluSer	730
Db	2341	CTGGATGCCAGCAGGGCCCCCGGAACTGTGCAACAGAGTCTCTCTCAGGAGTGAATCG	2400
Qy	731	LysProLysAspGluLeu 736	
Db	2401	AAGCCCAAGATGAGCTA 2418	
RESULT 10			
AAAD08507			
ID	AAAD08507	standard; cDNA; 2615 BP.	
XX			
AC	AAAD08507;		
XX			
DT	09-AUG-2001	(first entry)	
DE		Human secreted protein-encoding gene 2 cDNA clone HUVFV29, SEQ ID NO:30.	
KW		Human; secreted protein; proliferative disorder; cancer; tumour;	
KW		fetal abnormality; developmental abnormality; haematopoietic disorder;	
KW		immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;	
KW		inflammation; allergy; neurological disorder; Alzheimer's disease;	
KW		Parkinson's disease; cognitive disorder; schizophrenia; asthma;	
KW		skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;	
KW		cardiovascular disorder; angionenic disorder; kidney disorder;	
KW		gastrointestinal disorder; pregnancy-related disorder; gene therapy;	
KW		endocrine disorder; infection; wound healing; vulvexary; cell culture;	
KW		chemotaxis; food additive; binding partner identification; ss.	
OS	Homo sapiens.		
XX			
Key		Location/Qualifiers	
FH		24..545	
FT	CDS	/tag= a	
FT		/product= "Human secreted protein precursor"	
FT	sig_peptide	24..77	
FT		/tag= b	
FT	mat_peptide	78..542	



Qy 401 LysArgLeuGlnGlnLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
Db 1222 AAGAGATTGCAAGAGAAACAGAGTCAAGACGGGAAACAGCGTATCCCAAGGAG 1281  
Qy 421 IleGlyAsnLeuMetLysGlnLysGlnLysGlnLysValGluGlnLysThrLysGlnSerLeu 440  
Db 1282 ATTGGGAACCTTATGAGGAAATCGAGACCTTGTGGAAGAGAACCAAGAGAGTCACTG 1341  
Qy 441 AspValSerArgLeuThrArgGlnGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
Db 1342 GATGTGAGCAGATGATCCCGGGAGGTGGCCCTCTGCTATGAAGGCATCAGTCTCACC 1401  
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
Db 1402 ATGAACCTCCAACTCTCTGAATGTTCCAGCGGTGTGATGACGCGCTAATCTCTGAC 1461  
Qy 481 HisGlnCysGlnGlnLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
Db 1462 CAGAGTGTGAGAGCTGACAGACTGACCAATGTGACCACTTCCAGAGATGGCTAC 1521  
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGlnLysPheTyrGlyValThrValPheLys 520  
Db 1522 CGGGGTGAGACCTCCACATATCTCCCAATGAAAGTTCATGCTGCTACTGCTTCAAA 1581  
Qy 521 AlaLeuLysLeuGlnGlnGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
Db 1582 GCCTCTCAAGCTGGGCAAGGAAAGTTCCTGTCAGAGTGCCTCAGCTGCTACTACAAC 1641  
Qy 541 ValThrGlnLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
Db 1642 GTGACGGAGAGGTGGCGGCATCATGAGTCTCTCTCGCTGGATAGCGCCCTCTAC 1701  
Qy 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGlnGluValGlnAlaGluArgLys 580  
Db 1702 TTTTCTACTCTCATCTGGTGTGCGCACTGCCATCGCAAGAGTCCAGGAGAGAGAG 1761  
Qy 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
Db 1762 GATGATAGTATCATGATCCAGTGGACACTGATCTCTGATGCCAGACCTCTGTTGT 1821  
Qy 601 ValLysGluProProLafThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
Db 1822 GTCAAG 1881  
Qy 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
Db 1882 GACTTCGATCGCGAAACTTTTATTTTCACTGAACTGGATGCCAAGACCGTGAAGAGAG 1941  
Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
Db 1942 GTGACGCTCAGTGTGGAAG 2001  
Qy 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuAspPro 680  
Db 2002 GTGAAGGCTGTACCAAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2061  
Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMetLeuPheSerPro 700  
Db 2062 CGACACAG 2121  
Qy 701 GluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProGluPro 720  
Db 2122 GAAGAGATGACCTCTCCAG 2181  
Qy 721 AlaGlnLysLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2182 GCACAGAGAGTCTCTCAGGAGAGTGAATCGAAGCCCAAGAGATGAGCTA 2229

RESULT 11  
AAC64724  
ID AAC64724 standard; cDNA; 2829 BP.  
XX  
AC AAC64724;

XX 27-FEB-2001 (first entry)  
DT Human tumour suppressor Gros1-L encoding cDNA SEQ ID NO:1.  
DE  
DE Tumour suppressor; Gros1-L; Gros1-S; cell proliferation; regulation;  
KW cancer; cytostatic; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX WO2000065047-A1.  
XX  
PD 02-NOV-2000.  
XX  
PF 26-APR-2000; 2000WO-JP002731.  
XX  
PR 26-APR-1999; 99JP-00118806.  
XX  
PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
XX  
PI Wadhwa R, Sugihara T, Yoshida A;  
XX  
XX WPI; 2000-687340/67.  
DR P-PSDB; AAB36391.  
XX  
PT Full-length tumor suppressor genes encoding Gros1-L, Gros1-S and mouse  
PT homologs participating in regulation of cell proliferation, useful in  
PT development of preventives and remedies of cancer.  
XX  
PS Claim 1; Page 61-66; 114pp; Japanese.  
XX  
CC The present sequence encodes the human tumour suppressor designated Gros1  
CC -L. Gros1-L and Gros1-S have cytostatic activity and can be used in gene  
CC therapy. Gros1-L and Gros1-S genes are useful in the development of drugs  
CC used to treat and prevent cancer  
XX  
SQ Sequence 2829 BP; 659 A; 808 C; 801 G; 561 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 0 Length: 2829  
Score: 3780.00 Matches: 733  
Percent Similarity: 90.41% Conservative: 2  
Best Local Similarity: 90.16% Mismatches: 1  
Query Match: 97.67% Indels: 77  
DB: 3 Gaps: 2  
US-10-045-815-4 (1-736) x AAC64724 (1-2829)  
Qy 1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaAlaSer 20  
Db 52 ATGGCGGTACGGCGTTCAGAGCTGCTGACACACTGCTGGCTGCTGCGCGCTCC 111  
Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuPheAla 40  
Db 112 CAAGCCGAGGTGAGTCCGAGGAGGATGGGCGCATGGTGACCGCTGATCTCTCTGCC 171  
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspThrProGlyValValLeuSerMetGluArg 60  
Db 172 GAGGGGACCGAGACCTACGCGCGGGGACTGGCCCGGGTGGTCTCTGAGCATGAAAGG 231  
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
Db 232 GCGCTGCGCTCCCGGCGAGCCCTCGCGCCCTTGGCTGCGCTGCGCGCAGCAGTGTGCC 291  
Qy 81 AlaAspPheProThrGluLeuAspProAspThrProSerProAlaGlnAlaSerGly 100  
Db 292 GCGGACTTCCCGTGGAGCTGACCCCGACTGGTCCCCAGCCCGCCGAGCCCTCGGGC 351  
Qy 101 AlaGlyAlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgAlaAlaCysLeu 120  
Db 352 GCGCGCGCCCTCGCGAGCTGAGCTTCTTCGGGGGCTTCTGCGTGGCTGCTGCTGCTG 411  
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe 140



XX Human cDNA sequence SEQ ID NO:15159.  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
XX EF1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-00116126.  
XX 29-JUL-1999; 99JP-00248036.  
XX 27-AUG-1999; 99JP-00300253.  
XX 11-JAN-2000; 2000JP-00118776.  
XX 02-MAY-2000; 2000JP-00183767.  
XX 09-JUN-2000; 2000JP-00241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
XX Claim 8; SEQ ID NO 15159; 2537pp + Sequence Listing; English.  
XX The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention  
XX SQ Sequence 2993 BP; 666 A; 864 C; 856 G; 607 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 0 Length: 2993  
Score: 3726.50 Matches: 730  
Percent Similarity: 83.33% Conservative: 0  
Best Local Similarity: 83.33% Mismatches: 6  
Query Match: 96.29% Indels: 141  
DB: 4 Gaps: 1  
  
US-10-045-815-4 (1-736) x AAH16288 (1-2993)  
  
Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
Db 42 ATGGCGGTACCGGTTGAAGCTCTGACACACTGCTGGCTGTCGGCGCGTGCCTCC 101  
Qy 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40

Db 102 CAGCCGAGGTCGAGTCCGAGGAGGATGGGCGCATGGTGAGCGCTGATCTCTTCGCC 161  
Qy 41 GluGlyThrAlaAlaTyTrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
Db 162 GAGGGGACCGAGCCTACGCGCGGAGCTGGCGCGGGTGGTCTCTGAGCATGGAACGG 221  
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgCysArgThrGlnCysAla 80  
Db 222 GCGCTGCGCTCCCGGCGAGCCCTCGCGCCCTTCGCTGCGCTGCCGACCCAGATGTGCC 281  
Qy 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
Db 282 GCCGACTTCCGTCGGAGCTGGACCCGCTGCTCCCGCCAGCCGCGCCAGCCCTCGGGC 341  
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
Db 342 GCGCGCGCCCTCGCGACCTGAGCTTCTTCGGGGGCGCTTCGCGTGGCGCTGCCCTG 401  
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe 140  
Db 402 CGCGCTGCTCGCGCCCGCGCGCCGCTCGCTCAGCAGAGAGATGGAGTGGATTC 461  
Qy 141 ArgLysArgSerProTyTrAsnTyTrLeuGlnValAlaTyTrPheLysIleAsnLysLeuGlu 160  
Db 462 CGCAAGCGGAGCGCCCTCAACACTTACCTGCGAGGTGCGCTACTTCAAGATCAACAAGTTGGAG 521  
Qy 161 LysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetGln 180  
Db 522 AAAGTGTTCGTCAGCACACACCTTCTTCGTGGGCAATCTCTGAGCACATGGAAATGCGAG 581  
Qy 181 GlnAsnLeuAspTyTrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
Db 582 CAGAACCTAGACTATTACCAACCATGTCTGAGTGAAGGAGGCGGACTTCAAGGATCTT 641  
Qy 201 GluThrGlnProHisMetGlnLysPheArgLeuGlyValArgLeuTyTrSerGluGln 220  
Db 642 GAGACTCAACCCATATGCAAGATTTGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 701  
Qy 221 ProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTyTrPheValAlaTyTrGlu 240  
Db 702 CCACAGGAGCTGTGCCCCACCTAGAGGCGGCTGCAAGATATCTTTGSGCTATGAG 761  
Qy 241 GluCysArgAlaLeuGluGlyProTyTrAspTyTrAspGlyTyTrAsnTyTrLeuGluTy 260  
Db 762 GAGTGGCGTGGCTCTGCGAGGCGCTATGACTACGATGGCTACAACTACCTTGGATAC 821  
Qy 261 AsnAlaAspLeuPheGlnAlaThrThrAspHisTyTrIleGlnValLeuAsnCysLysGln 280  
Db 822 AACGCTGACCTCTCCAGGCCATCAGATCATTAATCCAGTCTCTCACTGTGAAGCAG 881  
Qy 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
Db 882 AACTGTGTCAAGGAGCTTGTTCCTCCCAAGTCCGAGAGAGCCCTTTGAAGACTTCCTC 941  
Qy 301 ProSerHisTyTrAsnTyTrLeuGlnPheAlaTyTrAsnIleGlyAsnTyTrThrGlnAla 320  
Db 942 CCATCGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1001  
Qy 321 GlyGluCysAlaLysThrTyTrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
Db 1002 GTTGATGTGCCAGAACCTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1061  
Qy 341 LeuAlaTyTrAlaAlaValMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
Db 1062 TTGGCTTATTATGCGAGCTATGCTTGGAGAAGAACACACACAGATCCATCGGCCCGCTGAG 1121  
Qy 361 SerAlaLysGluTyArgGlnArgSerLeuLeuLysGluLeuLeuPhePheAlaTyTr 380  
Db 1122 AGTGCACAGAGTACCGACAGCGAGCCCTACTTGGAAAAGAAAGAACTGCTTCTTCTGCTTAT 1181  
Qy 381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400



Db 1182 GATGTTTTTGGAAATCCCTTTTGGATCCGGAATTCATGGACTCCAGGAGAGATGATCC 1241  
Qy 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgLleSerGlnGlu 420  
Db 1242 AAGAGATTGCAGAGAAACAGAGATCAGAACGGGAAACAGCGGTACGCATCTCCAGGAG 1301  
Qy 421 IleGlyAsnLeuMetLysGluLleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
Db 1302 ATTGGGAACCTTATGAAGAAATCGAGACCTTGTGGAGAGAGAACCAAGAGTCACTG 1361  
Qy 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuThrGluGlyLleSerLeuThr 460  
Db 1362 GATGTGACAGACTGACCGGGAAGGTGGCCCTCTGTATGAAGGCATCAGTCTCACC 1421  
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrglnArgValValMetAspGlyValLleSerAsp 480  
Db 1422 ATGAACTCAAACTCTGAATGGTTCCAGCGGGTGTGTGATGGAGCGGTAAATCTCTGAC 1481  
Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaIleThrSerGlyAspGlyTy 500  
Db 1482 CACGAGTGTCAAGAGCTGCAGAGACTGACCAATGTGGCAGCAACTCAGGAGATGGTAC 1541  
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrglyValThrValPheLys 520  
Db 1542 CGGGTCAAGACTCCCACTACTCCCAATGAAAGTTCTATGGTGTCACTGTCTTCAA 1601  
Qy 521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyTyAsn 540  
Db 1602 GCCCTCAAGCTGGGCAAGAGCAAAAGTCTCTCGCAGAGTGCACCTGTACTACAAC 1661  
Qy 541 ValThrGluLysValArgAlaIleMetClnSerTyrglyPheArgLeuAspThrProLeuTy 560  
Db 1662 GTGACGGAAGAGTGGCGCATCATGGAGTCTTCCGCTCGGATACGCCCTCTAC 1721  
Qy 561 PheSerTySerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
Db 1722 TTTTCTACTCTATCTGTGTGGCAGTCCATCGAGAGGTCCAGGAGAGAGAG 1781  
Qy 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
Db 1782 GATGATAGTATCCAGTCCAGTGCAGCAACTGCATCTGTGATGCGGAGACCTCGTGTGT 1841  
Qy 601 ValLysGluProProAlaTyThrPheArgAspTyrglySerAlaIleLeuTyLeuAsnGly 620  
Db 1842 GTCAAAGAGCCCCAGCTTACCTTCGCGACTACAGCGCATCTTTTACCTTAATGGG 1901  
Qy 621 AspPheAspGlyGlyAsnPheTyThrGluLeuAspAlaTyThrValThrAlaGlu 640  
Db 1902 GACTTCGATGCGGAAACTTTATTTTCACTGAATGATGCAAGACCGTGAAGCGGAG 1961  
Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
Db 1962 GTGACGCTCAGTGTGAAGAGCGGTGGATCTCTTCAGCGACTGAAACCCACATGGA 2021  
Qy 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680  
Db 2022 GTGAAGGCTGTCAACAGGGGCGAGCGTGTGCCATCGCCCTGTGTTCACCTGGACCT 2081  
Qy 681 ArgHisSerGluArg----- 685  
Db 2082 CGACACAGCGAGCG- GGTGAGAGAGCTCGAGCGGGTGAAGAGAGCTGGTGTGTGTGTA 2140  
Qy 685 ----- 685  
Db 2141 CCCGTTCCAGAGCGCCTTGGTTGGCTTTCTTCTCCCAATCCATTCAGTGGCT 2200  
Qy 685 ----- 685  
Db 2201 GAGACAGAAAGAGACACTGGGACACCAGCTCCACGCCCTGTCTATTATGTCACATTG 2260  
Qy 685 ----- 685  
Db 2261 CCTTGTCTCCTGGGCTGTGTGAACGGGATCCAGGTGGGAAAGAGATCAAGACAGG 2320

Qy 685 ----- 685  
Db 2321 GAGCGATGCTGAGTTCTTGGTTCCCTCTTGGGCCCACTTCAGTGTCTTTCAGAG 2380  
Qy 685 ----- 685  
Db 2381 AGTAGGACCTGCTGGGAGGAGATGAGCCTGGGGCCATTAAAGAACCTTCCTTGTCCCT 2440  
Qy 685 ----- 685  
Db 2441 GGAAGTAGCAGCTGAGATAGCAGTGTCTGGAGCGGAGCCCTCTCTGAATGGGCAGG 2500  
Qy 686 ----- AspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
Db 2501 GGTGTGCTCTTCAGGACAGGGTGCAGGCAGATGACCTGGTGAAGATGCTCTTCAGCCCA 2560  
Qy 701 GluGluMetAspLeuSerGlnGluProLeuAspAlaGlnGlnGlyProProGluPro 720  
Db 2561 GAAGAGATGCTCTCTCCAGGAGCAGCCCTGGATGCCAGCAGGGGCCGCCCGAACCT 2620  
Qy 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2621 GCACAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2668

## RESULT 13

ABZ11345  
ID ABZ11345 standard; cDNA; 2152 BP.  
XX AC ABZ11345;  
XX XX  
DT 20-JAN-2003 (first entry)  
XX DE Human polynucleotide SEQ ID NO 227.  
XX KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
cell-proliferative disorder; neurodegenerative disease; bacterial;  
Parkinson's disease; Alzheimer's disease; autoimmune disease;  
multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
arthritis; cytostatic; immunomodulator; nootropic; dermatological;  
antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
antiarthritic; gene; ss.  
XX OS Homo sapiens.  
XX PN WO200270539-A2.  
XX PD 12-SEP-2002.  
XX XX  
XX PF 05-MAR-2002; 2002WO-US0005095.  
XX PR 05-MAR-2001; 2001US-00799451.  
XX (HYSE-) HYSEQ INC.

XX PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Ma Y, Yamaraki V, Chen R, Wang Z, Ghosh M,  
PI Wehrman T, Wang J, Wang D, Drmanac RT;  
XX WPI; 2002-759812/82.  
DR P-PSDB; ABP69128.  
XX PT New polynucleotides comprising sequences assembled from expressed  
sequence tags (ESTs), useful for treating cell-proliferative, or platelet  
neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or  
coagulation disorders.  
XX PS Claim 1; SEQ ID NO 227; 1012pp + Sequence Listing; English.

XX CC The invention relates to an isolated polynucleotide (I) comprising a  
nucleotide sequence selected from any of 948 sequences (ABZ11119-  
ABZ12066) or their mature protein coding portion, active domain coding



Alignment Scores:

Alignment Scores:

US-10-045-815-4 (

Ov 1 Met

—

CC	protein or	CC	Sequence 215	US-10-045-815-4 (	Qy	1 Met	32 ARG	21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC	identifying	CC	Alignment Scores:		Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC	encoded poly	CC	Pred. No.:		Db			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC	markers, as	CC	Score:		Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC	imaging, scr	CC	Percent Similarity		Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC	or proliferati	CC	Best Local Simila		Db			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC	or Alzheimer	CC	Query Match:		Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC	platelet or	CC	DB:		Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC	or lung fibr	CC			Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC	arthritis, e	CC			Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC	of the print	CC			Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC	directly fro	CC			Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC		CC			Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC		CC			Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC		CC			Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC		CC			Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC		CC			Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC		CC			Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC		CC			Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC		CC			Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC		CC			Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC		CC			Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC		CC			Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC		CC			Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC		CC			Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC		CC			Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC		CC			Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC		CC			Qy			21 Glu	61 Ala	81 Ala	101 Ala	121 Arg	141 Arg	161 Lys	181 Glu	201 Glu	221 Pro
CC		CC			Qy			2									

QY 601 ValLysGluProProAlaValThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
 DB 1742 GTCARAGAGCCCGCCAGCTACACCTTCGCGACTACAGCGCCATCTTACCTAATGGG 1801  
 QY 621 AspPheAspGlyGlyAspPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
 DB 1802 GACTTCGATGGCGGAACTTTTATTTCACTGAACCTGGATGCCAAGACCGTGCAGCGCAGAG 1861  
 QY 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
 DB 1862 GTGCGAGCTCAGTGTGGAAGAGCCGTGGGATCTCTTCAGCACTGAAACCCCATGGGA 1921  
 QY 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680  
 DB 1922 GTGAAGGCTGTCAACAGGGGCGAGCGCTGTGCCATCGCCCTGTGGTTCAACCTGGACCT 1981  
 QY 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
 DB 1982 CGACACAGCGCGGAGCAGGGTGCAGCAGATGACCTGTGTGAAGATGCTCTTCAGGCCCA 2041  
 QY 701 GluGlnMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProGluPro 720  
 DB 2042 GAAGAGATGGACCTCTCCAGAGCAGCGCCCTGGATGCCAGCAGGGGCCCGGACCT 2101  
 QY 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
 DB 2102 GCACAAGAGTCTCTCTCAGGCGAGTGAATCGAAGCCCAAGGATGAGCTA 2149

## RESULT 14

AAS17572  
 ID AAS17572 standard; cDNA; 2127 BP.

AC AAS17572;

DT 26-FEB-2002 (first entry)

DE DNA encoding novel secreted protein #1.

XX Secreted protein; cytostatic; immunosuppressive; vulnary; vaccine;  
 KW antiinflammatory; neuroprotective; nephrotropic; cardiovascular; human;  
 KW cancer; autoimmune disease; wound healing disorder; infection;  
 KW haematopoietic disorder; inflammatory disorder; infertility;  
 KW neurological disease; psychiatric disease; cardiovascular disease;  
 KW respiratory disease; renal; gastrointestinal; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT 1. .2127

FT CDS /tag= a

FT /product= "Human secreted protein"

XX WO200179454-A1.

XX 25-OCT-2001.

XX 11-APR-2001; 2001WO-US011797.

XX 13-APR-2000; 2000US-0196603P.

XX 24-APR-2000; 2000US-0199417P.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX (SMIK) SMITHKLINE BEECHAM PLC.

XX Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z;

XX WPI; 2002-061975/08.

XX P-PSDB; AAU09860.

XX New secreted proteins or polypeptides, useful for treating e.g. cancer,  
 XX autoimmune diseases, wound healing disorder, infections, hematopoietic  
 XX disorders, inflammatory disorders, infertility, cancer.

PS Claim 2; Page 34-35; 92pp; English.

XX The invention relates to an isolated novel secreted polypeptide (I) and  
 CC polynucleotide (II). (I) and (II) are useful for treating cancer,  
 CC autoimmune diseases, wound healing disorder, infections, hematopoietic  
 CC disorders, inflammatory disorders, infertility, neurological and  
 CC psychiatric diseases, cardiovascular diseases, respiratory diseases,  
 CC renal diseases, or gastrointestinal diseases. These may also be used to  
 CC treat diseases, abnormalities and disorders caused by abnormal  
 CC expression, production, function and/or metabolism of the genes, as  
 CC vaccines for inducing immunological response in a mammal, and in  
 CC screening methods for detecting the effect of added compounds on the  
 CC production of mRNA and polypeptide in cells. The polypeptides can be used  
 CC as immunogens to produce antibodies immunospecific for the polypeptides,  
 CC and to identify membrane-bound or soluble receptors. The polynucleotides  
 CC may be used as diagnostic reagents, in chromosome localisation studies,  
 CC and in tissue expression studies. The present sequence represents the  
 CC coding sequence of novel human secreted protein #1

XX SQ Sequence 2127 BP; 495 A; 623 C; 589 G; 420 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 0 Length: 2127  
 Score: 3572.00 Matches: 691  
 Percent Similarity: 94.57% Conservative: 5  
 Best Local Similarity: 93.89% Mismatches: 12  
 Query Match: 92.30% Indels: 28  
 DB: 6 Gaps: 3

US-10-045-815-4 (1-736) x AAS17572 (1-2127)

QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaAlaSer 20  
 DB 1 ATGGCGGTACGCGCGTTGAAGCTGCTGACACACACTGCTGGTGTGCTGGCGCTGCCCTCC 60  
 QY 21 GlnAlaGluValGluSerGluAlaGlyTyrPheGlyMetValThrProAspLeuLeuPheAla 40  
 DB 61 CAAGCGAGGTGCGAGTCCGAGCGAGATGGGCGATGGTACGCTGATCTGCTCTTCGCC 120  
 QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
 DB 121 GAGGGGACCGCAGCCTACGCGCGCGGAGCTGGCCCGGGTGGTCTCTGAGCATGGAACGG 180  
 QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
 DB 181 GCGTGGCTCCCGGGCAGCCCTCCGCGCCCTTCGCTGCGCTGCCGCCACCGATGGTCC 240  
 QY 81 AlaAspPheProTyrPheGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
 DB 241 GCCGACTTCCCGTGGGAGCTGGACCCCGACTGGTCCCCCAGCCCGCCAGGCTCGGGCC 300  
 QY 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
 DB 301 GCCCGCGCCCTGCGGAGCTGAGCTTCTCGGGGCGCTTCGCGTGGCGCTGCCCTCGCTG 360  
 QY 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
 DB 361 CGCGCTGCCCTCGGGCGCGCGCGCCCTCCCTGCGGCAATCTCTGAGCAGATGGAGTTC 420  
 QY 141 ArgLysArgSerProTyrAsnTyrIleuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
 DB 421 CGCAAGCGGAGCCCTACAACTACTCTGAGTCCGCTACTCTTCAAGATCAACAAGTTGAG 480  
 QY 161 LysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetGln 180  
 DB 481 AAAGCTGTGTGCGAGCACACACCTTCTTCGTGGCAATCTCTGAGCAGATGGAAATGCAG 540  
 QY 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
 DB 541 CAGAACCTAGACTATTACCAACCATGCTCTGAGTGAAGGAGGCGGAGCTTCAAGACTCT 600  
 QY 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGln 220

601	Db	GAGACTCAACCCATATGCAAGAAATTTCGACTGGGAGTCGGACTCTACTCAGAGGAAACAG	560
221	QY	ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGlnIuTyrPheValAlaTyrGlu	240
661	Db	CCACAGAAAGCTGTGTGCCACCTAGAGGGGGCGCTGCAAGAAATACTTTGTGGCCTATGAG	720
241	QY	GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr	260
721	Db	GAGTGGCGTGCCCTCTCGAAGGGCCCTATGACTACGATGGCTACAACTACCTTGTAGTAC	780
261	QY	AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln	280
781	Db	AAAGCTGACCTTTCAGGCCATCACAGATCATTACATCCAGTCTCTCAACTGTATGACAG	840
281	QY	AsnCysValThrGluLeuAlaSerHisProSerArgSLeuLysProPheGluAspPheLeu	300
841	Db	AACTGTGTCCAGGAGCTTGCCTTCCCAACCAAGTCGAGAAAGCCCTTTGAAGACTTCCCTC	900
301	QY	ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla	320
901	Db	CCATCCGATTAATTAATCTCGAGTTTGCTACTATAAC-----	939
321	QY	GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn	340
940	Db	-----AAGACA-----ATCTGCTATTGTAAT	960
341	QY	LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu	360
961	Db	CTTCTCTGTTCTCTGAATAATCTATAGAAA-----AAGAG	996
361	QY	SerAlaLysGlnTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaTyr	380
997	Db	AGTGCCCAAGGAGTACCGACAGCGAAGCTACTCGAAAAAGAACTGCTTTTCTTCGCTTAT	1056
381	QY	AspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIlePro	400
1057	Db	GATGTTTTTGGAAATCCCTTTTGGATTCGGATTTCATGGACTCCAGAAAGATGATCCC	1116
401	QY	LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu	420
1117	Db	AAGAGATTTCAGAGAGAAACAGAACTCAGAACGGGAAACAGCCGTAGCATCTCCACAGAG	1176
421	QY	IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu	440
1177	Db	ATTGGGAACCTTATGAAGGAATTCGAGACCTTTGTGGAGAGAAAGCAAGGAGTCACTG	1236
441	QY	AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr	460
1237	Db	GATGTGACGAGACTGACCCGGGAAGTGGCCCTGCTGTATGAGGCATCAGTCTCACC	1296
461	QY	MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp	480
1297	Db	ATGAACCTCCAAACTCTCTGAATGTTTCCACGGGTGGTGATGACGCGGTAATCTCTGAC	1356
481	QY	HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaIleThrSerGlyAspGlyTyr	500
1357	Db	CACGAGTGTACAGAGCTGACAGACTGACCAATGTGGCAGCAACCTCAGAGATGGCTAC	1416
501	QY	ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys	520
1417	Db	CGGGGTACAGACCTCCCCACATACTCCCAATGAAAGTTCTATGGTGTCTACTGCTTCAA	1476
521	QY	AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn	540
1477	Db	GCCTCAAGCTGGGGCAAGAGGCAATGTTCTCTGACAGTGCCTACCTGTACTACAAC	1536
541	QY	ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr	560
1537	Db	GTACAGAGAAAGTGGCGGCATCATGAGTCTCTACTTCCGCTTGGNATCGCCCCCTAC	1596
561	QY	PheSerTyrSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArgLys	580
1597	Db	TTTTTCTACTCTCATCTGGTGTGCGCACTGCCATCGAAGAGTTCAGGCAGAGAGAAG	1656

RESUIT 15

AAC64726

ID AAC64726 standard; cDNA; 2416 BP.

AAC64726;

Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100
1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	

DI 27-FEB-2001 (LISC emily/ XX

Mouse tumour suppressor Grosl-L encoded

XX

KW Tumour suppress

KW cancer; cyto

**XXIX**

OS  
xy  
Mus musculus.

PN WO2000065047-A1.




PD 02-NOV-2000.

XX

PF 26-APR-2000; 2000WO-JP002731.

XX  
PP 75-1000-1000 007B-00118806

XX  
10000000-0000 10000000-0000

PA (CHUG-) CHUGAI RES INST MOLECULAR MED

XX  
-----  
XX

PI Wadhwa R, Sugihara T, Yoshida A;

[illegible]

DR WPI; 2000-68/340/67.  
DR D DCDR. AAB25302

XX  
SQ Sequence 2416 BP; 579 A; 671 C; 683 G; 482 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.:	1.56e-298	Length:	2416
Score:	3323.00	Matches:	630
Percent Similarity:	92.37%	Conservative:	36
Best Local Similarity:	87.38%	Mismatches:	52
Query Match:	85.87%	Indels:	4
DB:	3	Gaps:	2

US-10-045-815-4 (1-736) x AAC64726 (1-2416)

QY	11	ThrLeuLeuAlaValValAlaAlaSer---GlnAlaGluValGluSerGluAlaGly	29
DB	40	ACGATGCTAGCGGTTCGCGCGCGCGCGCGCTTACGGGTTCGGAC-GAGTCTGAGCGCGGA	98
QY	30	TrpGlyMetValThrProAspLeuLeuPheAlaGluGlyThrAlaAlaTyAlaArgGly	49
DB	99	TGGGACGTGGCAGCCCTGACCTGTTTACGAGAGGGACCGCGGCTTACTCGCCAGG	158
QY	50	AspTppProGlyValValLeuSerMetGluArgAlaLeuArgSerArgAlaAlaLeuArg	69
DB	159	GACTGGCCGGGTGGTCTCTGAACATGAGCGGCTCTGCGCTCGCGGCGGCGCTCGCT	218
QY	70	AlaLeuArgLeuArgCysArgThrGlnCysAlaAlaAspPheProTrpGluLeuAspPro	89
DB	219	GGCCTCGGCTCGCTGCGCACACCTGTGCCACCACTGCGCGGGGACCGGACCTG	278
QY	90	AspTppSerProSerPro-----AlaGlnAlaSerGlyAlaGlyAlaLeuArgAspLeu	107
DB	279	GATCTCGGTCCGACCCAGCCTGAGCCAGGACCGCGCGCGCGCTGACGACCTG	338
QY	108	SerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeuArgArgCysLeuGlyProPro	127
DB	339	CGCTCTTTCGGAGCGGTGCTCGCGGTGCCCTTGCCTAGCGCGCTGCTCGGCGCGCC	398
QY	128	AlaAlaHisSerLeuSerGluGluMetGluLeuGluPheArgLysArgSerProTyrAsn	147
DB	399	TC TGCCCACTCTGCTGAGTGAGAACTGGACCTGGAGTTCAACAGCGGAGCCCGCTACAC	458
QY	148	TyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGluLysAlaValAlaAlaHis	167
DB	459	TACCTGCAGTTCGCTATTTCAAGATAAAACAAGCTGAGAAAGCTGTGGCTGGCGCAC	518
QY	168	ThrPhePheValGlyAsnProGluHisMetGluMetGlnGlnAsnLeuAspTyrTyrGln	187
DB	519	ACCTTCTTTGTGGGCATTCCTGAGCATTGAGATGGGCAGAACTTCGACTATTACC	578
QY	188	ThrMetSerGlyValLysGluAlaAspPheLysAspLeuGluThrGlnProHisMetGln	207
DB	579	ACCATGTTCTGGGTGAAGGAGCGACATTCAGGGATCTCGAGGCCAAGCCCATATGCAT	638
QY	208	GluPheArgLeuGlyValArgLeuTyrSerGluGluGlnProGlnGluAlaValProHis	227
DB	639	GAGTTTCGGCTGGGGGTACGACTCTACTCAGAGAGAGAACCCAGAAAGCTGTGCCCA	698
QY	228	LeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGluGluCysArgAlaLeuCysGlu	247
DB	699	CTGGAGGCGGCACGTCAAGAGTACTTTGTGGCCGATGAGGAGTGGCTGCCCTCTCGAA	758
QY	248	GlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAla	267
DB	759	GGGCGCCTATGACTACGAGCGGTACAACCTACTAGACTACAGCGCTGACCTCTTCAGG	818
QY	268	IleThrAspHisTyrIleGlnValLeuAsnCysLysGlnAsnCysValThrGluLeuAla	287
DB	819	ATCACAGATCATTAGCTCCAGTCTCTCACTGTATAGCAGAACTGTGTCCAGAGTGGCT	878
QY	288	SerHisProSerArgGluLysProPheGluAspPheLeuProSerHisTyrAsnTyrLeu	307
DB	879	TCCACCAACGATAGGAAACCCCTTTGAGACTTCTCCCTTCACTATTAATTAACCTA	938

Db	2019	CACGGCTGGCCATCGCCCTGTGTTTCACGCTGGATCCTCGGCACAGTGAGAGACAGG	2078
Qy	688	valGlnAlaAspLeuValLysMetLeuPheSerProGluGluMetAspLeuSerGln	707
Db	2079	GTCAGGCAGATGACCTGGTGAAGATGCTGTTTCAGCCCAAGAGAGGTGGACCTCCCCCAG	2138
Qy	708	GluGlnProLeuAspAlaGlnGlnGlyProProGluProAlaGlnGluSerLeuSerGly	727
Db	2139	GAACAGCCCTGCTGACCAGCAGGGTTCGCCAGAGCCTGGAGAGAGTTTCTGCATGCT	2198
Qy	728	Ser	728
Db	2199	GCT	2201

Search completed: July 18, 2004, 10:20:02  
Job time : 1086 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 08:21:37 ; Search time 53 Seconds  
(without alignments)  
723.087 Million cell updates/sec

Title: US-10-045-815-4

Perfect score: 3870

Sequence: 1 MAVRAKLKLTLLAVVAAS.....PPEPAQSLSGSEKPKDEL 736

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	572	14.8	401	1	CASP HUMAN
2	536.5	13.9	400	1	CASP MOUSE
3	501	12.9	437	1	NO55 HUMAN
4	500.5	12.9	431	1	SC65 RAT
5	419	10.8	271	1	CASP CHICK
6	157.5	4.1	603	1	UVR_C_CHLPN
7	124	3.2	1129	1	EG27_CABEL
8	120.5	3.1	2291	1	SPCB DROME
9	118.5	3.1	609	1	CEP2 MOUSE
10	118	3.0	496	1	MOQ3 STAEF
11	118	3.0	1597	1	CTRO MOUSE
12	117	3.0	1286	1	CTRO HUMAN
13	116	3.0	884	1	RA50 SULSO
14	116	3.0	1186	1	ADDB BACSU
15	115.5	3.0	1969	1	MYSA CABEL
16	114.5	3.0	613	1	SG2 BOVIN
17	114	2.9	928	1	NIBA HUMAN
18	114	2.9	3911	1	AKA9 HUMAN
19	113	2.9	1127	1	V855 TREPA
20	113	2.9	1805	1	HMW2 MYCGE
21	112.5	2.9	1726	1	FBPA AQUAE
22	112.5	2.9	1893	1	CSD2 HUMAN
23	112	2.9	887	1	NED4 MOUSE
24	112	2.9	1694	1	CLH DICI
25	111	2.9	2431	1	POLN SFV
26	110.5	2.9	879	1	RA50 SULTO
27	110.5	2.9	1257	1	CCAA BACTU
28	110	2.8	388	1	CUS5 HUMAN
29	110	2.8	859	1	IFTS AQUAE
30	109.5	2.8	543	1	MEFS HUMAN
31	109	2.8	1000	1	NED4 HUMAN
32	109	2.8	2230	1	GOA4 HUMAN
33	108.5	2.8	887	1	NED4 RAT

34	108	2.8	1084	1	MYSS RABIT
35	108	2.8	2663	1	CENE HUMAN
36	107.5	2.8	624	1	BPBD BACSU
37	107.5	2.8	714	1	FRDA_HELPY
38	107	2.8	605	1	UVR_C_CHLGV
39	106.5	2.8	349	1	CUS5 MOUSE
40	106	2.7	503	1	EGLX MOUSE
41	106	2.7	705	1	YK22 YEAST
42	106	2.7	1109	1	MYSD DICI
43	105	2.7	693	1	EX70 DROME
44	105	2.7	3210	1	CENF HUMAN
45	105	2.7	4687	1	PLEI_RAT

ALIGNMENTS

RESULT 1

CASP\_HUMAN

ID CASP\_HUMAN STANDARD; PRT; 401 AA.

AC O75718;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Cartilage-associated protein precursor.

GN CRTAP OR CASP.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Fetal brain;

RX MEDLINE=20169181; PubMed=10702664;

RA Tonachini L., Morello R., Monticone M., Skaug J., Scherer S.W.,

RA Cancedda R., Castagnola P.,

RT "cDNA cloning, characterization and chromosome mapping of the gene

encoding human cartilage associated protein (CRTAP).";

RL Cytogenet. Cell Genet. 87:191-194(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettner M., Maman A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grumman J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,

RT "Generation and initial analysis of more than 15,000 full-length

human and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -!- SUBCELLULAR LOCATION: Secreted. Extracellular matrix (By

similarity).

CC -!- TISSUE SPECIFICITY: Found in articular chondrocytes. Expressed in

CC a variety of tissues.

CC -!- SIMILARITY: BELONGS TO THE CRTAP / NO55 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR ENBL; AJ006470; CAA07054.1; -;  
DR ENBL; BC008745; AA008745.1; -;  
DR Genew; HGNC:2379; CRTAP.  
DR MIM; 605497; -;  
DR InterPro; IPR008940; Prenyl trans.  
DR Extracellular matrix; Signal.  
KW Extracellular matrix; Signal.  
FT SIGNAL 1 26 POTENTIAL.  
FT CHAIN 27 401 CARTILAGE-ASSOCIATED PROTEIN.  
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 363 363 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 137 137 E -> D (IN REF. 2).  
SQ SEQUENCE 401 AA; 46561 MW; 4BEED4089195456F CRC64;  
Query Match 14.8%; Score 572; DB 1; Length 401;  
Best Local Similarity 34.4%; Pred. No. 9.5e-30;  
Matches 145; Conservative 68; Mismatches 168; Indels 40; Gaps 11;  
QY 4 RALKLLTLLAVAA--ASQAEVSEAGMGVTPDLL-----FAEGTAAYARGDMPGVVL 56  
DB 6 RGAALLALLCVACALRAGRAQYERYSFRRDELMPLESAYRHLDKYSGEHMAESVG 65  
QY 57 SMERALSRAALRALRLRCCTCAADFPWELDPWSPSPAQASGAGALRDLSPFGGLLR 116  
DB 66 YLEISLHLRLDSEAFCHRNCSA-----APQPEPAAGLASYPBLRFGGLLR 115  
QY 117 AACLRRC-LGPPAAHSL--SEEMELEFRKSPYNLYQVAYFKINKLEKAAVAATTFVGN 173  
DB 116 AHCLKRCCKGLPAFRQSPREVLADQREPKYKFLQFAVFKANLPKAAIAAHTLLKH 175  
QY 174 PEHMEMQNLDYQTMGSVKAEADFKDLETPHMQEFLGVRLYSEQPOQAVPHLEAALQ 233  
DB 176 PDDEMMQRNMYKSLFGAEDY-IKDLETKSYSLFIRAVYNGENWRTSITDMELALP 234  
QY 234 EYFVAYECCALCGPYDYGYNLYEYNADLFOAITDHYIOVLNCKQNCVTELASHPSRE 293  
DB 235 DFFKAFVECLAACGSEIKDKF-----DFYLSIADHYVEVLECKIQEENL-TPVIGG 287  
QY 294 KPPEDFLPSPHYNLYQFAYNIGNYTOGCECAKTYLLFFPNDVMMQNLYAYAM-----L 348  
DB 288 YPVEKFAVATMYHYLQFAYKLNLDKNAAPCAVSYLLFDQNDKVMQNLYVYQVHRDTWGL 347  
QY 349 GEETHRSIGRESKEYRQSLLEKELLFFAYDVGIFPFVDPDSWTPPEVPIKRLQEKOK 408  
DB 348 SDEHFQ---PRPEAVQFNVTTLQKELYDPAKE-----NIMDDGEVEVYDLDLELET 400  
QY 409 S 409  
DB 401 S 401  
RESULT 2  
CASP MOUSE STANDARD; PRT; 400 AA.  
ID -CASP MOUSE  
AC Q9CYD3; O88698;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Cartilage-associated protein precursor.  
GN CRTAP OR CASP.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=99357019; PubMed=10429950;  
RA Morello R., Tonachini L., Monticone M., Viggiano L., Rocchi M.,  
RA Cancedda R., Castagnola P.;

RT "cDNA cloning, characterization and chromosome mapping of Crtap  
RT encoding the mouse cartilage associated protein.";  
RN Matrix Biol. 18:319-324(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Embryo;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikolaio I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix. Expressed in  
CC -!- TISSUE SPECIFICITY: Found in articular chondrocytes. Expressed in  
CC a variety of tissues.  
CC -!- SIMILARITY: BELONGS TO THE CRTAP / NO55 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC ENBL; AJ006469; CAA07053.1; -;  
DR ENBL; AK017797; BAB30938.1; -;  
DR MGD; MGI:1891221; Crtap.  
DR InterPro; IPR008940; Prenyl trans.  
KW Extracellular matrix; Signal.  
FT SIGNAL 1 25 POTENTIAL.  
FT CHAIN 26 400 CARTILAGE-ASSOCIATED PROTEIN.  
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 362 362 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 7 7 A -> T (IN REF. 2).  
FT CONFLICT 34 34 N -> S (IN REF. 2).  
FT CONFLICT 79 80 SE -> RQ (IN REF. 2).  
FT CONFLICT 249 250 SR -> VA (IN REF. 2).  
FT CONFLICT 254 254 D -> T (IN REF. 2).  
SQ SEQUENCE 400 AA; 46166 MW; 812245C6A4E769E2 CRC64;  
Query Match 13.9%; Score 536.5; DB 1; Length 400;  
Best Local Similarity 34.9%; Pred. No. 1.9e-27;  
Matches 130; Conservative 64; Mismatches 145; Indels 33; Gaps 10;  
QY 46 YARGDMPGVLSMERALRSRAALRALRLRCCTCAADFPWELDPWSPSPAQASGAGALR 105  
DB 54 YSGEHWAESVGLVLSLRLHLRLDSEAFCHRNCSAATP-----APAPA 103  
QY 106 DLSPFGGLLRRAACLRC-LGPPAAHSL--SEEMELEFRKSPYNLYQVAYFKINKLEKA 162  
DB 104 ELRLFGSVLRRACQLKRCQGLPAFRQSPRSVLADFPQQREPYKFLQFAVFKANDLPKA 163  
QY 163 VAAATTFVGNPEHEMNMQNLDYQTMGSVKAEADFKDLETPHMQEFLGVRLYSEQPO 222  
DB 164 IAAATYLLRHPDDMMKRNMYEYKSLPGA-EDHKLETKSYSLFIRAVYNGENWNR 222  
QY 223 EAVPHEAALQEFYFVAYEECCALCEGPDYDGYNLYEYNADLFOAITDHYIQVLCNKQNC 282



```

Db 223 TSISDMELALPDFELKAFYECLACESREIKDFK-----DFYLSADHYVEVELEKIR 276
Qy 283 VTSLASHPSREKPEFPLFSHYNYLOFAYNYIGNYTOAGCAKTYLLFPFNDEVMNQNL 342
Db 277 -EETLTPVIGYPVEKVFVATWYHLQFAYYKLNLDKNAAPCAVSYLLFPQSDRVMOQNLV 335
Qy 343 YYAAM-----LGEHRTSIPRESAKEYRQBSLLEKELFFAYVDVFGIFVDPDSWTPEE 397
Db 336 YYQVHRDKWGLSDHFQ---PRPEAVQFNVTLQKELYDFQE-----HLMDDDEGEVVE 388
Qy 398 VIPKRLQEKOKS 409
Db 389 YVDDLLETESA 400

RESULT 3
NO55 HUMAN STANDARD; PRT; 437 AA.
AC Q92791; Q9H4F6;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Nucleolar autoantigen NO55.
GN SC65 OR NOL55.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bladder epithelium;
RX MEDLINE=97015880; PubMed=8862517;
RA Ochs R.L., Stein T.W. Jr., Chan E.K.L., Ruutu M., Tan E.M.;
RT "cDNA cloning and characterization of a novel nucleolar protein.";
RL Mol. Biol. Cell 7:1015-1024 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=2041121; PubMed=10952778;
RA Fossa A., Siebert R., Aasheim H.C., Maelandsmo G.M., Berner A.,
RA Fossa S.D., Paus E., Smeland E.B., Gaudernack G.;
RT "Identification of nucleolar protein No55 as a tumour-associated
autoantigen in patients with prostate cancer.";
RL Br. J. Cancer 83:743-749 (2000).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Moquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -I- SUBCELLULAR LOCATION: Nuclear; nucleolar. Localized uniformly
throughout the granular component of the nucleolus and on the
surface of chromosomes during mitosis.
CC -I- SIMILARITY: BELONGS TO THE CRTAP / NOS5 FAMILY.
CC
```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U47621; AAC51792.1; --
DR EMBL; AJ250583; CAC16786.1; --
DR EMBL; BC001047; AAH01047.1; --
DR EMBL; BC007942; AAH07942.1; --
DR EMBL; BC011701; AAH11701.1; --
DR GK; Q92791; --
DR GO; GO:000230; C:nuclear mitotic chromosome; TAS.
DR GO; GO:0005730; C:nucleolus; TAS.
DR GO; GO:0005716; C:synaptonemal complex; TAS.
DR GO; GO:0007130; P:synaptonemal complex formation; TAS.
DR InterPro; IPR008940; Prenyl_trans.
KW Nuclear protein; Antigen.
FT DOMAIN 346 437 ASP/GLU-RICH (ACIDIC).
FT CONFLICT 186 186 Q -> R (IN REF. 2).
SQ SEQUENCE 437 AA; 50381 MW; 50C82FCE9BB7274A CRC64;

Query Match 12.9%; Score 501; DB 1; Length 437;
Best Local Similarity 33.3%; Pred. No. 4,1e-25;
Matches 138; Conservative 53; Mismatches 191; Indels 32; Gaps 10;

Qy 12 LLAVVAASQAQEVSEAGCGWVTPDLL-----PAEGTAAVARGDWPGVVLSEALRSRA 66
Db 8 LLWLLGSAQAQYKYSRGPFPDDLPLAAAYGHAELOVEGESWRESARVLEALRLHR 67
Qy 67 ALRALRLRCRTQCAADFF-WELDPDWSFSPAQAGALRDLSPFGGLLRRAACLRCLG 125
Db 68 LLRDSEAFCHANGSGPAAPKPDGGRADWAC-----ELRLFGRLERAACLRCKR 121
Qy 126 PPAHSL---SEEMLEPRKSPYNYLOVAYFKINKLEKAYAAAHFPVGNPEHMEMQON 182
Db 122 TLPAQVPPYPPRQLRDFQSLPYQYLHYALFKANRLEKAVAAAYTFLQRPKHETAKY 181
Qy 183 LDYQTMGSGKEADFKLETPHMQEPRGLGVRLYSEEQPAQVPHLEALQOYFVAYEEC 242
Db 182 LNYQGLMDVADESITDLEAQYEAFLRAVKNLNSGDFRSTEDMERALSAYLAVPARC 241
Qy 243 RALCEGPDYDGYNYLYNADLFOAITHYIQVNLKQNCQVTELASHPSRKPFD-FLP 301
Db 242 LAGCEGAHEQVDFK-----DFYPAIALFAESLQCKVDCENLT--PNVGIFYVDKFA 293
Qy 302 SHYNYLOFAYNYIGNYTOAGCAKTYLLFPFNDEVMNQNLAY--AAMLGEHRTSISGP 358
Db 294 TMYHLYQAYYKLNLDVROAARSAASYMLDFDSDVMQNLVYFHRARWGLEB-EDPQP 352
Qy 359 RESAKYRQSRSLLEKELFFAYVDVFGIFVDPDSWTPEEVIKRLQEKOKSERE 412
Db 353 REEAMLYHNQTAELRELLEFTH-----MYLQSDDEWELEETEPPELPEDALSDAE 402

RESULT 4
SC65 RAT STANDARD; PRT; 431 AA.
AC Q64375;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Synaptonemal complex protein SC65.
GN SC65.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
CC
```

RX MEDLINE=93213429; PubMed=1363622;  
RA Chen Q., Pearlman R.E., Moens P.B.;  
RT "Isolation and characterization of a cDNA encoding a synaptonemal  
RT complex protein."  
RL Biochem. Cell Biol. 70:1030-1038(1992).  
CC !- SUBCELLULAR LOCATION: Nuclear. Located in the pairing zone of the  
CC synaptonemal complex.  
CC !- TISSUE SPECIFICITY: Found in testis, brain, heart and at a much  
CC lower level in liver.  
CC !- SIMILARITY: BELONGS TO THE CRTAP / NOS5 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X65454; CAA46449.1; -.  
DR PIR; A56822; A56822.  
DR InterPro; IPI008940; Prenyl\_trans.  
KW Nuclear protein.  
FT DOMAIN 319 328 GLU-RICH (ACIDIC).  
FT DOMAIN 319 384 GLU-RICH (ACIDIC).  
FT SEQUENCE 431 AA; 49995 MW; 4A34F3029407B2E7 CRC64;  
SQ  
Query Match 12.9%; Score 500.5; DB 1; Length 431;  
Best Local Similarity 35.9%; Pred. No. 4.4e-25;  
Matches 124; Conservative 47; Mismatches 153; Indels 21; Gaps 8;  
Qy 46 YARDGPGVLSMERALRSRAALRLRCRTCAADFPWELDPDWSPPS- --AQSAGAG 102  
Db 14 YEGSWRESARYLEAALRLHLLRDSFAFCHANGCSG- PATSPRPAPGPDGNEGDDG 71  
Qy 103 ALRDLSPFGLLRRAACLRCLGPPAHS- --SEMELEFRKRSYNYILOVAFKINKL 159  
Db 72 WARELRFGHVLRAACLRCKTLPAPQYPSRQLRDFONRLPYQYLYAHAFKANRL 131  
Qy 160 EKAVAAHTTFVGNPEHMEQONLDYQTMGSKVEADFKDLETQPMQBFRLGVRLYSEE 219  
Db 132 EKAVAAHTTFQRNPKELAKYLYNYRGLMDIGDESITLDEAQPYEAVFLQAVKLYNSG 191  
Qy 220 QPQAVPHLEALQOYFVAYBECCALCEGPDYDGVNLYNADLFOAITDHYIQVNLCK 279  
Db 192 DFRSSTHEMERALADYMTVFARCLAGCEGAHQVDFK- ----DFYPAIADUFAESLOCK 245  
Qy 280 QNCVTELASPSREKPED- FLPSHYNLOFAYNIGNYTOAGECAKTYLLPPNDENVN 338  
Db 246 VDCEANLT- PNVGGFFVDKFAVTWYHLYLOFAYKLVNDVQARSASMYLDFPKDSVMQ 303  
Qy 339 QNLAY- --AAMIGEHTSRIGPSAKERYQRSLLEKELLFFAY 380  
Db 304 QNLVYFRHRAWGLEE- EDFQPREAVLYHNTSELRELLDFT 347  
RESULT 5  
ID CASP\_CHICK STANDARD; PRT; 271 AA.  
AC Q90830;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Cartilage associated protein precursor (Dualin).  
GN CRTAP OR CASP.  
OS Gallus Gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97360293; PubMed=9217321;  
RA Castagnola P., Gennari M., Morello R., Tonachini L., Marin O.,  
Gaggero A., Cancedda R.;  
RT "Cartilage associated protein (CASP) is a novel developmentally  
RT regulated chick embryo protein."  
RL J. Cell Sci. 110:1351-1359(1997).  
CC !- SUBCELLULAR LOCATION: Secreted; extracellular matrix.  
CC !- TISSUE SPECIFICITY: Found in articular chondrocytes. Expressed in  
CC a variety of tissues.  
CC !- SIMILARITY: BELONGS TO THE CRTAP / NOS5 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X97607; CAA66206.1; -.  
DR Extracellular matrix; Signal.  
FT SIGNAL 1 15 POTENTIAL.  
FT CHAIN 16 271 CARTILAGE ASSOCIATED PROTEIN.  
FT CARBOHYD 76 76 N-LINKED (GLCNAC...) (POTENTIAL).  
FT SEQUENCE 271 AA; 31493 MW; A56343DBF604914C CRC64;  
SQ  
Query Match 10.8%; Score 419; DB 1; Length 271;  
Best Local Similarity 36.1%; Pred. No. 4.2e-20;  
Matches 100; Conservative 46; Mismatches 107; Indels 24; Gaps 7;  
Qy 11 TLLAVVAASQAQVESEACGNGWTPDLL- ----FAEGTAAYARGDWGVLSMERALRSR 65  
Db 4 TLLAALLATAGAQVERYSFRSPRDELMPLLESAYRGLDQYSTENWPESVSYLEVSRLY 63  
Qy 66 AALRALRLRCRTCAADFPWELDPDWSPPSQAASGAGALRDLSPFGLLRRAACLRRC-L 124  
Db 64 RLLRDTAEAFCHNCSS- ----AGPLTAPPAD- ----GELAEALLAGVLRRAQCRLRCKQ 114  
Qy 125 GPAAHLSUSEMEL- --EPRKRSYNYILOVAFKINKLEKAVAAHTTFVGNPEHMEQON 182  
Db 115 GLPAPRAQPGRELEEFQRREPKYLOFAYKANNLPKAIKAAHTTFLKHPDDMMQON 174  
Qy 183 LDYQTMGSKVEADFKDLETQPMQBFRLGVRLYSEEQPOEAVPHLEALQOYFVAYEEC 242  
Db 175 MAYKSIPTAE- HIKDLETKPYENLFVRAVRAINGDNWRTSISDMELALPDPFFKTYDDC 233  
Qy 243 RALCEGPDYDGVNLYNADLFOAITDHYIQVNLCK 279  
Db 234 IAACEGSRERIKDFK- ----DFYLSIADHYIEVLACK 264  
RESULT 6  
ID UVRC\_CHLPN STANDARD; PRT; 603 AA.  
AC Q926W6; Q9JQA6;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE UVrABC system protein C (Uvrc protein) (Excinuclease ABC subunit C).  
GN UVRC OR CPN0940 OR CP0921 OR CP0974.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydiales.  
OX NCBI\_TaxID=833558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CNL029;  
RX MEDLINE=99206606; PubMed=10192388;  
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hymen R.W.,  
Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."  
RL Nat. Genet. 21:385-389(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=AR39;

RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Winn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.,  
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 RT pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406 (2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.,  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CWL029 from USA.";  
 RL Nucleic Acids Res. 28:2311-2314 (2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TW-183;  
 RA Geng M.M., Schumacher A., Muehldorfer I., Bensch K.W., Schaefer K.P.,  
 RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.,  
 RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with  
 RT other Chlamydia strains based on whole genome sequence analysis.";  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: The UvrABC repair system catalyzes the recognition and  
 CC processing of DNA lesions. UvrC both incises the 5' and 3' sides  
 CC of the lesion. The N-terminal half is responsible for the 3'  
 CC incision and the C-terminal half is responsible for the 5'  
 CC incision (By similarity).  
 CC -!- SUBUNIT: Interacts with uvrB in an incision complex (By  
 CC similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to the uvrC family.  
 CC -!- SIMILARITY: Contains 1 UVR domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC  
 DR EMBL; AE001673; AAD19078.1; -.  
 DR EMBL; AE002251; AAF38706.1; -.  
 DR EMBL; AP002548; BAA99148.1; -.  
 DR EMBL; AE017160; AAP98903.1; -.  
 DR PIR; B86608; B86608.  
 DR PIR; F72017; F72017.  
 DR TIGR; CP0921; -.  
 DR HAMAP; MF 00203; -; 1.  
 DR InterPro; IPR001943; UvrB/C.  
 DR InterPro; IPR004791; UvrC.  
 DR InterPro; IPR001162; UvrC.C.  
 DR InterPro; IPR00305; UvrC.N.  
 DR Pfam; PF01541; Excl\_endo\_N; 1.  
 DR ProDom; PD005870; UvrC; 1.  
 DR SMART; SM00465; Glyc; 1.  
 DR TIGRFAMs; TIGR00194; uvrC; 1.  
 DR PROSITE; PS50151; UVR; 1.  
 DR PROSITE; PS50164; UVR; 1.  
 DR PROSITE; PS50165; UVR; 2; 1.  
 KW SOS response; Excision nuclease; DNA repair; DNA recombination;  
 KW DNA excision; Complete proteome.  
 FT DOMAIN 205 240 UVR.  
 SQ SEQUENCE 603 AA; 69426 MW; 28017E4FA19FFC84 CRC64;

Query Match 4.1%; Score 157.5; DB 1; Length 603;  
 Best Local Similarity 20.98; Pred. No. 0.01; 223; Indels 193; Gaps 27;  
 Matches 131; Conservative 81; Mismatches 193;

QY 51 WPGVLSMERALRSRAALRALRL-----RCRTQCAADFPWELDPWSPSPAQASG 100  
 DB 111 WPKV-----EAIKTAITSSQRLIFGYPVSAEACHILL-----EVSQWFP----- 152  
 QY 101 AGALRDLISFFGGLLRAAC-----LRRCLGPPAAHSLSEMELEFRKSPNYLQVAFKI 156  
 DB 153 ---LRTCSDFEALRKPCILYDMKRLCAPCVGCTPEEYOGTLDK-----AILFLK 201  
 QY 157 NKLEKAVAAAHFTFFGNPEHMEMQONLDYYOTMGSVKAEADPKDLETOPHMQEF----- 209  
 DB 202 KIEEVVKDLKVIQKASDNLEFEQAANYTTLISLIQKAMAKQVEXFHFONIDALGIYR 261  
 QY 210 -----RLGVRLYS-ESQQAQAVPHLSAALQEFVA--YECCALCEGYP 250  
 DB 262 HKQRTILTLTVRSQKILGARHFFSFFENAQDQDLSLFIQYVVSQPIPKKILTLPL 321  
 QY 251 DYDGVNLYENAD---LFOAITDHYIQVL-----NCKQNCVTELASHPSEKPFEDF-- 299  
 DB 322 EFTLSIV-LNAESPRLRSPTGIGKELDLATRNKAYAAATL---PSTLTPYQDFON 377  
 QY 300 --LPSHNYLOFAYYNIQNTQAGECAKTYLLFFPFNDVNMQNLAYYAAMLGEBHTRSIG 357  
 DB 378 ILRMSQYPIRECYDNA--HMQGAHATGVIVFENNGFDPKQ-----YRTFSI- 423  
 QY 358 PRESAREYRQSLLEKELLFPAYDVFGIPFVDPDSWTPPEVVPKELQEKQKSERETAVRI 417  
 DB 424 --DSEKTDNLALLEEVLL-----RRFHSLLTALPDMIV---VDGGKTHYNNKTKKI 469  
 QY 418 SOEIGNLMKETLVEEKTES--LDVSRLTREGPLLYEGISLTMSKLLNGYQVRVMD 475  
 DB 470 IQLNLTGIOVTTAKESNHSRGLNKEKIFCEIFP---EGFSLPPTSNNLLOFFQ----- 521  
 QY 476 GVISDHECOELQRLTNVAATSGDGYRGOTSPHTNEKPYGVTVFKALKLQGEQKVPLOSA 535  
 DB 522 -ILRD-----EAHRFAISKH---RKRGKALFEQEKIPGIGEV----- 555  
 QY 536 HLYNVNTEKVRIMESYFRLDTPLVFSYSHLVCTAIEVQAEKDDSHPVVDNCILNA 595  
 DB 556 -----KRRLLQK-----FKSWKQVMSQSELEA----- 580  
 QY 596 ETLVCVKEPPAYTRDYSAILYLNGDFD 623  
 DB 581 -----IPGLTKKDIALLARQKDFN 600  
 RESULT 7  
 EG27\_CABEL STANDARD; PRT; 1129 AA.  
 AC Q09228; Q09229; Q8MOP3; Q9XYD0;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Egg laying defective protein 27.  
 GN EGL-27 OR C04A2.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM A), FUNCTION, SUBCELLULAR LOCATION, AND  
 RP DEVELOPMENTAL STAGE.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99128194; PubMed=9927605;  
 RA Herman M.A., Ch'ng Q., Hattenbach S.M., Ratliff T.M., Kenyon C.,  
 RA Herman R.K.;  
 RT "EGL-27 is similar to a metastasis-associated factor and controls cell  
 RT polarity and cell migration in C. elegans.";  
 RL Development 126:1055-1064 (1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Du Z.;  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.



RX MEDLINE=92335263; PubMed=16311106;  
RA Byers T.J., Brandin E., Lue R., Winograd E., Branton D.;  
RT "The complete sequence of Drosophila beta-spectrin reveals  
RT supra-motifs comprising eight 106-residue segments.";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:6187-6191(1992).  
RN [2]  
RP SEQUENCE OF 1-800 FROM N.A.  
RX MEDLINE=90009037; PubMed=2677025;  
RA Goldstein L.S.B.;  
RT "Sequence similarity of the amino-terminal domain of Drosophila beta  
RT spectrin to alpha actinin and dystrophin.";  
RL J. Cell Biol. 109:1633-1641(1989).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Branton R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Bokorva D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Flosser C., Gabrielson A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei V., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapieton M., Strong R., Sun E.,  
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
RN [4]  
RP CHARACTERIZATION.  
RX MEDLINE=88059242; PubMed=3680372;  
RA Dubreuil R., Byers T.J., Branton D., Goldstein L.S.B., Kiehart D.P.;  
RT "Drosophila spectrin. I. Characterization of the purified protein.";  
RL J. Cell Biol. 105:2093-2102(1987).  
RN [5]  
RP STRUCTURE BY NMR OF 2145-2262.  
RX MEDLINE=96164435; PubMed=8591029;  
RA Zhang P., Talluri S., Deng H., Branton D., Wagner G.;  
RT "Solution structure of the pleckstrin homology domain of Drosophila  
RT beta-spectrin.";  
RL Structure 3:1185-1195(1995).  
CC -!- FUNCTION: Spectrin is the major constituent of the cytoskeletal  
CC network underlying the erythrocyte plasma membrane. It associates  
CC with band 4.1 and actin to form the cytoskeletal superstructure of  
CC the erythrocyte plasma membrane. Interacts with calmodulin in a  
CC calcium-dependent manner.

CC -!- SUBUNIT: Native spectrin molecule is a tetramer composed of two  
CC antiparallel heterodimers joined head to head so that each end  
CC of the native molecule includes the C-terminus of the alpha  
CC subunit and the N-terminus of the beta subunit.  
CC -!- SIMILARITY: Belongs to the spectrin family.  
CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.  
CC -!- SIMILARITY: Contains 1 PH domain.  
CC -!- SIMILARITY: Contains 17 spectrin repeats.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M92288; AAA28399.1; --  
CC EMBL; AE003506; AAF48751.1; --  
CC PIR; A46147; A46147.  
CC PDB; 1DKO; 03-APR-96.  
CC Flybase; FBgn0003471; beta-Spec.  
CC GO; GO:0045169; C:fusome; IDA.  
CC GO; GO:0005886; C:plasma membrane; IDA.  
CC GO; GO:0045170; C:spectrosome; IDA.  
CC GO; GO:0008017; F:microtubule binding; IDA.  
CC GO; GO:0007274; P:neuromuscular synaptic transmission; IMP.  
CC InterPro; IPR001589; Actbind actnin.  
CC InterPro; IPR001715; Calponin-like.  
CC InterPro; IPR001849; PH.  
CC InterPro; IPR002017; Spectrin.  
CC InterPro; IPR001605; Spectrin\_PH.  
CC Pfam; PF00307; CH; 2.  
CC Pfam; PF00169; PH; 1.  
CC Pfam; PF00435; spectrin; 17.  
CC PRINTS; PR00683; SPECTRINPH.  
CC SMART; SM00033; CH; 2.  
CC SMART; SM00233; PH; 1.  
CC SMART; SM00150; SPEC; 17.  
CC PROSITE; PS00019; ACTININ\_1; 1.  
CC PROSITE; PS00020; ACTININ\_2; 1.  
CC PROSITE; PS50021; CH; 2.  
CC PROSITE; PS50003; PH DOMAIN; 1.  
CC Cytoskeleton; Membrane; Calmodulin-binding; Actin-binding; Repeat;  
CC 3D-structure; Actin capping.  
FT DOMAIN 1 271 ACTIN-BINDING (BY SIMILARITY).  
FT DOMAIN 50 154 CH 1.  
FT DOMAIN 169 271 CH 2.  
FT REPEAT 298 408 SPECTRIN 1.  
FT REPEAT 418 522 SPECTRIN 2.  
FT REPEAT 524 633 SPECTRIN 3.  
FT REPEAT 635 739 SPECTRIN 4.  
FT REPEAT 741 844 SPECTRIN 5.  
FT REPEAT 846 950 SPECTRIN 6.  
FT REPEAT 952 1057 SPECTRIN 7.  
FT REPEAT 1059 1167 SPECTRIN 8.  
FT REPEAT 1169 1273 SPECTRIN 9.  
FT REPEAT 1275 1378 SPECTRIN 10.  
FT REPEAT 1380 1485 SPECTRIN 11.  
FT REPEAT 1487 1591 SPECTRIN 12.  
FT REPEAT 1593 1697 SPECTRIN 13.  
FT REPEAT 1699 1804 SPECTRIN 14.  
FT REPEAT 1806 1910 SPECTRIN 15.  
FT REPEAT 1912 2016 SPECTRIN 16.  
FT REPEAT 2018 2078 SPECTRIN 17.  
FT DOMAIN 2147 2259 PH.  
FT CONFLICT 2278 2278 D -> Y (IN REF. 2).  
FT STRAND 2150 2157  
FT TURN 2166 2167  
FT STRAND 2173 2179  
FT STRAND 2185 2187  
FT HELIX 2192 2195  
FT TURN 2197 2198



[illegible]





```
QY 354 -----RSIGPRE-----SAKEYR-----QRSLEKELLFPAYOV 382
Db 334 RVRELQMLDTEKOSRARADQRTESQVVELAVKEHKAELIALQOALKEQKL----- 386
QY 383 FGIPFVDPDSWTE-----EVIPKRLQEKOKSERETAVRISQIGNLMKEIET- 430
Db 387 -----KAESLSDKLNDLEKKHAMLEWNAESLQOKLETERELKQRLLEQAKLQOQMDLQ 440
QY 431 -----LVEETKESLDVSRUTR-EGGPLYEGISLWNSKLNGYQVVDGVISDHCO 484
Db 441 KNHIFRLTQLOALDRADLLKTERSDEYQ-----LENIQVLYSHEKVRMEGTISO----- 492
QY 485 ELQRLTNVAATSGDVGRTSPHTPNEKFGVTVFVKALQGO-----RGKVPLOSAHLYNV 541
Db 493 -----QTK-----LIDFLOAKMDQPAKKKVPLOQNELKAL 524
QY 542 TEKVRIMESYFLDTPLYFSYSHLVCRTAIEVQAEKDDSH 585
Db 525 EKEKARCAE-----LEAL--QKTRIELRGAREEA-AHRKATDHP 561

RESULT 13
RASO_SULSO
ID RASO_SULSO STANDARD; PRT; 864 AA.
AC Q97WH0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR SSO2249.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awayez M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Crauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -!- SIMILARITY: Belongs to the SMC family. RAD50 subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AEO06829; AAK42417.1; -.
CC PIR; B90395; B90395.
CC HAMAP; MF_00449; -.
CC InterPro; IPR003439; ABC transporter.
CC InterPro; IPR007517; Rad50_zn_hook.
CC InterPro; IPR003405; SMC_C.
CC InterPro; IPR003395; SMC_N.
CC Pfam; PF04423; Rad50_zn_hook; 1.
CC Pfam; PF02483; SMC_C; 1.
```

```
DR Pfam; PF02463; SMC_N; 1.
DR ProDom; PD00006; ABC transporter; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 30 37 ATP (BY SIMILARITY).
FT DOMAIN 164 701 COILED COIL (POTENTIAL).
SQ SEQUENCE 864 AA; 101601 MW; 657076AAEA9B709FC CRC64;

Query Match
Best Local Similarity 18.3%; Pred. No. 8.1;
Matches 128; Conservative 103; Mismatches 227; Indels 242; Gaps 29;

QY 141 RKRSY-----NYLOVAVFKINKLEKAVAAAHATFFVGNPEHMEMQONLDYYQTMGSVKREAD 196
Db 232 KRNQYKLTTTLTKIKGEGELNARSI-----EELRKQTEMDQLE----- 272
QY 197 FKDLTQPMQEFRLGVRLY-----SEEQPOEAVPHLEAALQYFVAYEECRALCEGYD 251
Db 273 -KEINELNLRNLIKLPKPEYEVLAKSHTMSANVINLEKEIEEYKAIKRRKEEL---EPK 328
QY 252 YDGYNYLEYNADLFOAITDHYIQV-----LNCKQNC---VTELASH---PSREKFFE 297
Db 329 YLYKLELERKLEELQPKYQYKYLKLSLDLSKLNKERLEKDASELSNDIDKVNSELEQKVE 388
QY 298 DFLPSHYNYLQFAYVNYIGNYTOAGECAKTYLLFFPNDEVMN-----ONLAYYAALMGE 350
Db 389 ETRKKQLNL-----RAQLAKVESLISEKNEIINNISOVEGETCPVGRPLDE 435
QY 351 EHTRSIGPRESAKYRORSLEKLEKLLFFAYDVFGIPFVDPDSWTEPEVIPKRLQEKOKSE 410
Db 436 EHKQKI--IKEAKSYILQLELNKEL-----EBELKKITNELNKIE 474
QY 411 RE-----TAVRISOEIGNLMKEITL--VVEK----- 435
Db 475 REYRLSNKASYDNVWRQLKKNELLENLHSELESKNDEEEKKINEEVKELKLYEE 534
QY 436 -----TKESLDVSRLTREGGLLYEGISLTM-----NSKLLN-GYQRV 472
Db 535 FMRLSKYTKBELDKKRVKLDKMKKKEIEKEMRGLESELKGLDRKALESKILDENKRV 594
QY 473 VMD-----GVISH-----ECQELQLTNVAATSGDGYRGQSPHTPNEKFGVGT 517
Db 595 KLDKEMKKKGILEDIYIRQVKLLQBEVKNLREEVNIIQFDENRY----- 637
QY 518 VFKALKLQGEKVPLOSAHLYNVYTEKVRIMESYFLDTPLYFSYSHLVCRTAIEEV-- 575
Db 638 -----NELKTSIDA-----YNLSLKEKENRKS--RIEGELE-----SLEKDIENISN 677
QY 576 -----QAEKDDSHPVHVDNLCILNAETLVCVCKEPPATFRDYSAILYN-----GDFDGG 625
Db 678 RIANYELQLKDREXIIINAINKLEKIRGALGERKLQSYIIMTTKOLIENLNLDIISKFD-- 735
QY 626 NFYFTELDKATVTAEVQPOCGRA-----VGFSSGTENPHGVKAVTRGQRCALAL--- 674
Db 736 -----LSIKKVNEMIPKTRGGRSSSGDILVYNSGDTLP--IVSLSGGERIALSLALR 787
QY 675 -----WFTLDRHSERDRVQADDLVKMLFSPPEE 702
Db 788 LAIAKALMSNTNFFILDEPTIHLDDQKAYLIEIRAAKE 827

RESULT 14
ADDB_BACSU
ID ADDB_BACSU STANDARD; PRT; 1166 AA.
AC P23477;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE ATP-dependent nuclease subunit B.
GN ADDB OR Bsu10620.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
```

RP SEQUENCE FROM N.A.  
RC STRAIN=OGL;  
RX MEDLINE=91267926; PubMed=1646786;  
RA Kooistra J., Venema G.;  
RT "Cloning, sequencing, and expression of *Bacillus subtilis* genes  
RT involved in ATP-dependent nucleic acid synthesis";  
RL J. Bacteriol. 173:3644-3655(1991).  
[2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=168;  
RX MEDLINE=98240224; PubMed=9579061;  
RA Noback M.A., Holsappel S., Kiewiet R., Terpstra P., Wambutt R.,  
RA Wedler H., Venema G., Bron S.;  
RT "The 172 kb *prkA*-*addA* region from 83 degrees to 97 degrees of the  
RT *Bacillus subtilis* chromosome contains several dysfunctional genes,  
RT the *glyB* marker, many genes encoding transporter proteins, and the  
RT ubiquitous hit gene";  
RL Microbiology 144:859-875(1998).  
[3]  
RN SEQUENCE FROM N.A.  
RP STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerthoff I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
RA Ertan K.D., Errington J., Fabret C., Fertari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Ghit S., Glaser P., Coffeau A., Golightly E.J., Grandi G.,  
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.P., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetle D., Porwollik S., Prescott A.M.,  
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler H., Wedler H., Weitzengger T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*";  
RL Nature 390:249-256(1997).  
CC -!- FUNCTION: THE ENZYME COMPLEX MAY HAVE A WIDE VARIETY OF CATALYTIC  
CC ACTIVITIES INCLUDING ATP-DEPENDENT EXONUCLEASE, ATP-STIMULATED  
CC ENDONUCLEASE, ATP-DEPENDENT HELICASE AND DNA-DEPENDENT ATPASE  
CC ACTIVITIES.  
CC -!- SUBUNIT: THE B SUBUNIT IS ATP-DEPENDENT NUCLEASE COMPLEX IS FORMED  
CC BY ONE B SUBUNIT IN CONJUNCTION WITH ONE A SUBUNIT.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed, usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M63489; AAA22200.1; -;  
CC EMBL; Y14081; CAA74481.1; -;  
CC EMBL; Z99109; CAB12902.1; -;  
CC PIR; A39432; A39432.  
CC Subtilist; BG10465; addB.

DR InterPro; IPR000212; UvrD-helicase.  
DR Pfam; PF00580; UvrD-helicase; 1.  
KW Hydrolase; Nuclease; Exonuclease; Endonuclease; Helicase; ATP-binding;  
KW DNA repair; Complete proteome.  
FT NP\_BIND 8 15 ATP (POTENTIAL).  
SQ SEQUENCE 1166 AA; 134631 MW; 1A57BBE81A08AB4A CRC64;  
Query Match 3.0%; Score 116; DB 1; Length 1166;  
Best Local Similarity 19.1%; Pred. No. 12;  
Matches 134; Conservative 95; Mismatches 230; Indels 244; Gaps 33;  
QY 127 PAASLSSEMELEERKSPYNYLVQVAFKINKLEKAVAAHAAFTFVGNGPEHMEQQLDY- 185  
DB PSYREPEHELEL-FRMTGK-----TYRLHQKAK-----ELNLDIT 274  
QY 186 YQTMG-----VKEADFKDLETPHMQBFLGVRILYSBEQOEAVPHLEA- 231  
DB YKELSGTERHTKTPELAHLAQ---YEARPAIP-YAEK--QEALTVMQANRRAELEGIA 328  
QY 232 -----LQEFYVAYEECPALCEGPDYDGY---NYLEYNADLFOAITDHYIQLNCKQNC 282  
DB REIHALVREKGYRYKDVAILARQEPEDYKMDYKVFADYEIPYF---IDGKASMLN----- 380  
QY 283 VTELASHPSREKPFEDFLPSHYNYLVQVAFKINKLEKAVAAHAAFTFVGNGPEHMEQQLDY- 333  
DB -----HPLIEFIRSSLDVLK-----GNWRYEAVFRCVKTLLPLNEPKAKVR 423  
QY 334 ---DEVNQNLAAYVAMLGEEHRSIGRESAKY----- 365  
DB 424 EQVDLENYCIAY--GIKGRWTK--GDRFOYRRFVSLDDDDFAQTQDEIEMENMLNDRD 479  
QY 366 -----RQSLLEKELLFFAYDVFGIPFVDPDSWTSWTEEVIPKRL-CEKOKSE 410  
DB 480 WIVPLFOLQKMKKAKIVQKAEALRY-----LEETDVLKLDQERQAE 526  
QY 411 RETAVRISQEIKNLMEITLVE-----KTKESLVSRLTREGGPLLVEGILSTNWSK 464  
DB 527 DDGRIIEAQHQQAWDVAILLEBFVEMMGDEISLDLFOQMIEAGA---ESLTFSLIPP 583  
QY 485 LLNGQVRVMDGVSDHCEQBLTLNVAATSGYRGQTSPTHPNE-----K 512  
DB 584 AL-----DQVFVGN--MOLSRMYGTSCTFVLGANDGVLPARPDENGVLSDDRWLK 633  
QY 513 FYGVTVFKALKLGQEGKVPLOSAHLYNVTEKVRIMESYERLDTF---LYFSY--SHLV 567  
DB 634 TIGV-----ELSSGGRRLLEHFLI-----YMAFSSPSDRLYVSYPIADAE 675  
QY 568 CHTAIEEVOAERKDDSHPVHVNDICILNAETLVCVKEPPAYTFRDYSAILYNGDFDGNF 627  
DB 676 GNTLLPSMIVKRLLEELFPHHKEILLTN-----EPEQVS--DEBQMLVYVNVKSVQSP 725  
QY 628 YTELDKATVTAETVQPCQGRAVFGSSGTENPHGVKAVTRGQRCALALWFTLLDPRHSRDR 687  
DB 726 TASQL-----RLMTREYDISDVWVWSTVYVLMSEQDR 756  
QY 688 VQADDLVKMLFSPPEMDLSQOPIDAQGGPPEPAQESLSGSES 730  
DB 757 LQSKKLFSLSFRNE--VKQLERSVSRQLYGRIQGSVSRMET 797

## RESULT 15

MYSA\_CABEL STANDARD; PRT; 1969 AA.  
ID MYSA\_CABEL  
AC P12844; Q21440;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Myosin heavy chain A (YHC A).  
DE MIO-3 OR K12F2.1.  
OS Caenorhabditis elegans.  
GN Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Meloiderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;

[1]  
SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=89178677; PubMed=2926820;  
RA Dibb N.J., Maruyama I.N., Krause M., Karn J.;  
RT "Sequence analysis of the complete Caenorhabditis elegans myosin  
heavy chain gene family.";  
RL J. Mol. Biol. 205:603-613(1999).  
[2]  
SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Harris B.R.;  
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: Muscle contraction.  
CC -1- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2  
heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)  
and 2 regulatory light chain subunits (MLC-2).  
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
CC -1- DOMAIN: The rodlike tail sequence is highly repetitive, showing  
cycles of a 28-residue repeat pattern composed of 4 heptapeptides,  
characteristic for alpha-helical coiled coils.  
CC -1- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light  
meromyosin (LMW) and 1 heavy meromyosin (HMM). It can later be  
split further into 2 globular subfragments (S1) and 1 rod-shaped  
subfragment (S2).  
CC -1- MISCELLANEOUS: There are four different myosin heavy chains in  
C. elegans.  
CC -1- MISCELLANEOUS: MHC A and MHC B are found exclusively in the body  
wall muscle. They co-assemble into body wall thick filament.  
CC -1- SIMILARITY: Contains 1 myosin-like globular head domain.  
CC -1- SIMILARITY: Contains 1 IQ domain.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; X08067; CAA30856.1; -;  
EMBL; Z78193; CAB01576.1; -;  
PIR; T23622; S02771.  
DR HSSP; P08799; 1MND.  
DR WormPep; K12F2.1; CE12204.  
DR InterPro; IPR000048; IQ region.  
DR InterPro; IPR001609; myosin head.  
DR InterPro; IPR004009; Myosin\_N.  
DR InterPro; IPR002928; Myosin\_tail.  
DR Pfam; PF00612; IQ; 1.  
DR Pfam; PF00063; myosin head; 1.  
DR Pfam; PF02736; Myosin\_N; 1.  
DR Pfam; PF01576; Myosin\_tail; 1.  
DR PRINTS; PR00193; MYOSTINHEAVY.  
DR ProDom; PD000355; myosin\_head; 1.  
DR SMART; SM00015; IQ; 1.  
DR SMART; SM00242; MYSC; 1.  
DR PROSITE; PSS0096; IQ; 1.  
DR ATP-binding; Muscle protein; Coiled coil; Thick filament; Actin-binding;  
ATP-binding; Multigene family.  
FT DOMAIN 1 793  
FT DOMAIN 794 823  
FT DOMAIN 857 1969  
FT NP\_BIND 179 186  
FT DOMAIN 667 689  
FT DOMAIN 770 784  
FT MOD\_RES 130 130  
FT CONFLICT 116 116  
FT Y -> YVRKLLKPKKIKINILNFQ (IN REF.  
2).  
SQ SEQUENCE 1969 AA; 225509 MW; 64577BBAP7EAD80A CRC64;

Query Match 3.0%; Score 115.5; DB 1; Length 1969;  
Best Local Similarity 21.7%; Pred. No. 28;

	Matches	76;	Conservative	54;	Mismatches	134;	Indels	87;	Gaps	13;
QY	155	KINKLEKAVAAAH	FFVGNPEHMEQ	QNLDYYQTM	SGVKEAD	FKDLETQPHMQE	FRILGVR	214		
Db	920	KINQLKATLESK	LSDIITGOLED	MQ-ERNED	LARQK--KKT	DQSLSDTKKHVQ	DLELSLR	975		
QY	215	LYSEEPQOEAV	PHLEAALQEF	VAYEECRAL	CEGPDYD	GVNLYEYNAD	LFOAITD--H	271		
Db	976	--KAEQEKQSR	DHINIRSLQD	EMANQDEAV	AKNKEKHQ	EBESNRKLNED	L-QSBE	KVNH	1032	
QY	272	YIQVLNCKQNC	VTELAGHPS	REK---P	FEFLPSH	NYNYLQFAY	NYNYTQAGE	CAKTY	327	
Db	1033	LEKIRNKLEQ	QMDLEENID	REKRSRG	DIKAKK	VEGDLKVAQ	ENIDEITKQ	KHDVETT	1092	
QY	328	L-----L	FFNDEV	MNQN-----	LAYVAM	LGEHTR	SGIPRESA	KEYQRS	369	
Db	1093	LKRKEEDL	HHTNAKLA	ENNSIIAK	LQRLIK	ELTARNAE	LEEELEA	ERNRSQ--K	SDRSRS	1150
QY	370	LLEKELLFF	AYDVFG	IPFVDP	SDSWTP	BEVIPKRLQ	E-----	K	406	
Db	1151	EAEREL-----	EELTER	LEQOQG	GATAAQL	EANKKEA	EIAKLRR	1189		
QY	407	QKSE-----	RETAV-----	RISQ	IGNLMKE	ETITLVE	EKTESLDV	SRITRE	448	
Db	1190	EKEEDSLN	HETAIS	SLRKRG	HGDSVA	ELTEQLE	TQLKLKAK	SEAEKSK	LQRD	1240

Search completed: July 18, 2004, 09:41:14  
Job time : 57 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 17:57:06 ; Search time 6546 Seconds  
(without alignments)  
3357.557 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 736  
Sequence: 1 MAVRALKLLTLLAVVAAS.....PPEPAOESLSGSEKPKDEL 736

Scoring table:  
OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Word size: 1

Total number of hits satisfying chosen parameters: 55025477

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2/1/USPTO.spool\_P/US10045815/runat\_14072004\_123119\_17251/app\_query.fasta\_1.903  
-DB=EST -QPMF=fastap -SUFFIX=oli.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USRP=US10045815@cgn.1.1.6425@runat\_14072004\_123119\_17251 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WAKN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estto:\*  
8: em\_hrc:\*  
9: gb\_estl:\*  
10: gb\_estc:\*  
11: gb\_hrc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_man:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vri:\*  
28: gb\_gsei:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	359	48.8	2583	11	BC004160	BC004160 Homo sapi
2	245	33.3	848	13	BUI77586	BUI77586 AGENCOURT
3	244	33.2	905	13	BUI47415	BUI47415 AGENCOURT
4	240	32.6	929	13	BQ677232	BQ677232 AGENCOURT
5	237	32.2	899	13	BUI19488	BUI19488 AGENCOURT
6	230	31.2	884	13	BUI190877	BUI190877 AGENCOURT
7	194	26.4	937	10	BF206567	BF206567 601870908
8	192	26.1	989	13	BQ678884	BQ678884 AGENCOURT
9	190	25.8	705	10	BE294207	BE294207 601172737
10	187	25.4	885	13	BQ680853	BQ680853 AGENCOURT
11	185	25.1	870	13	BQ676738	BQ676738 AGENCOURT
12	181	24.6	899	13	BUI148566	BUI148566 AGENCOURT
13	178	24.2	881	13	BUI543499	BUI543499 AGENCOURT
14	174	23.6	955	13	BQ678346	BQ678346 AGENCOURT
15	168	22.8	730	10	AM964170	AM964170 EST376243
16	165	22.4	675	13	EX109119	EX109119 BX109119
17	163	22.4	2211	29	AY411090	AY411090 Homo sapi
18	163	22.1	575	9	AW068699	AW068699 cn21h12.x
19	162	22.0	823	13	BUI59350	BUI59350 AGENCOURT
20	160	21.7	913	12	BG679334	BG679334 602628511
21	160	21.7	941	13	BUI38872	BUI38872 AGENCOURT
22	155	21.1	883	12	BG681584	BG681584 602628071
23	153	20.8	840	12	B1114871	B1114871 602861453
24	152	20.7	555	12	BG983273	BG983273 IL5-CN006
25	152	20.7	1108	12	BM473346	BM473346 AGENCOURT
26	151	20.5	1101	13	EX364036	EX364036 BX364036
27	148	20.1	981	13	BUI39422	BUI39422 AGENCOURT
28	148	20.1	1201	13	EX334235	EX334235 BX334235
29	147	20.0	592	14	CB369351	CB369351 T9BST2V98
30	147	20.0	633	10	AM966813	AM966813 EST378887
31	147	20.0	867	14	CF265106	CF265106 AGENCOURT
32	147	20.0	999	14	CA488534	CA488534 AGENCOURT
33	145	19.7	926	9	AL521774	AL521774 AL521774
34	144	19.6	594	10	BF343590	BF343590 602014425
35	143	19.4	668	10	BE298108	BE298108 601118250
C 36	143	19.4	1041	13	EX364035	EX364035 BX364035
37	140	19.0	651	14	CB055054	CB055054 NISC_SM07
C 38	140	19.0	1201	13	EX337092	EX337092 BX337092
39	138	18.8	528	9	AU280069	AU280069 AU280069
40	137	18.6	957	12	BG420284	BG420284 602448302
41	133	18.1	766	12	BI856252	BI856252 603382985
42	133	18.1	959	12	BM802225	BM802225 AGENCOURT
C 43	132	17.9	686	14	CB055053	CB055053 NISC_SM07
C 44	128	17.4	434	10	AM838215	AM838215 QV2-LT005
45	128	17.4	777	12	BG120604	BG120604 602346739

ALIGNMENTS

RESULT 1	BC004160	2583 bp	mRNA	linear	HTC 19-NOV-2003
LOCUS	Homo sapiens leucine proline-enriched proteoglycan (leprecan) 1,				
DEFINITION	mRNA (CDNA clone IMAGE:2824480), containing frame-shift errors.				
ACCESSION	BC004160				
VERSION	BC004160.2	GI:37588922			
KEYWORDS	HTC.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 2583)				

## AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L.H., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Kettman, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Burtner, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E., Scherch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257  
12477932  
2 (bases 1 to 2583)  
Strausberg, R.

Direct Submission  
Submitted (01-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
On Oct 8, 2003 this sequence version replaced gi:14708765.  
Contact: MGC help desk  
Email: [ccapbs-remail.nih.gov](mailto:ccapbs-remail.nih.gov)  
Tissue Procurement: DCTD/DTF  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
contact: [amadansystemsbiology.org](mailto:amadansystemsbiology.org)  
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IPAL Plate: 2 Row: h Column: 23  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21361917  
This clone has the following problem: frame shifted.

## FEATURES

Location/Qualifiers  
source  
1..2583  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2824480"  
/tissue\_type="Lung, small cell carcinoma"  
/clone\_id="NIH MGC 7"  
/lab\_host="DH10B-R"  
/note="vector: pOTB7"

## ORIGIN

## Alignment Scores:

Pred. No.: 0 Length: 2583  
Score: 359.00 Matches: 732  
Percent Similarity: 99.05% Conservative: 0  
Best Local Similarity: 99.05% Mismatches: 4  
Query Match: 48.76% Indels: 7  
DB: 11 Gaps: 0

US-10-045-815-4 (1-736) x BC004160 (1-2583)

1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaSer 20

Db QY 32 ATGGGGTACGCGCGCTTGAAGCTGCTGACCACTGCTGGTGTGGCGCGCTGCTCC 91  
QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40  
Db 92 CAAGCGAGGTGAGTCCGAGCGAGATGGGCGATGGTGGCGCTGATCTGCTTCGCC 151  
QY 41 GluGlyThrAlaAlaValAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
Db 152 GAGGGGACCGCAGCCTACGCGCGCGGACCTGGCCCGGGTGGTCTCTGAGCATGGAACGG 211  
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlyCysAla 80  
Db 212 GCGCTGCGCTCCCGGCGAGCCCTCGCGCCCTTCGCTCGCGTCCGCCACCATGTGCC 271  
QY 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
Db 272 GCGGACTTCCCGTGGAGCTGACCGCCGACCTGGTCCCGAGCCCGCGCCCTCGGCC 331  
QY 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLe 120  
Db 332 GCGGC-CGCCCTGCGCGACCTGAGCTTCTCGGGGCGCTTCTGCGTCCGCTGCGCT 390  
QY 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140  
Db 391 GCGCGCTGCTCGGCGCGCGCGCCCACTCGCTCAGCGAAGAGATGGAGCTGAGTT 450  
QY 140 eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGl 160  
Db 451 CCGCAAGCGGAGCCCTACACTACCTCGAGTCCGCTTCAAGATCAACAGTTGGA 510  
QY 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGl 180  
Db 511 GAAAGCTGTGCTGCGAGCAGCACACCTTCTTCTGGGCAATCTCTGAGCACATGGAATGCA 570  
QY 180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
Db 571 CGAGAACCTAGACTATTACCAACCATCTCTCGAGTGAAGGAGCGCGACTTCAAGGATCT 630  
QY 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGl 220  
Db 631 TGAGACTCAACCCCATATGCAAGAAATTCGACTGGAGTGCAGCTTCTACTCAGAGGAACA 690  
QY 220 nProGlnAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGl 240  
Db 691 GCCACAGGAAGCTGTGCGCCCACTAGAGGCGCGCTGCAAGATACCTTGTGGCTATGA 750  
QY 240 uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTy 260  
Db 751 GAGTGGCGTGGCTCTCGAAGGCGCTATGACTACGATGGTACAACTACCTTGAAGA 810  
QY 260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGl 280  
Db 811 CACCGCTGACCTTCTCCAGGCCATCACAGATCATTACATCCAGTCTCTCACTGAAGCA 870  
QY 280 nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
Db 871 GAACGTGTCTACGGAGCTTGTCTCCCAAGTCGAGAGAGCCCTTTGAAGACTTCTCT 930  
QY 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320  
Db 931 CCCATCGCATTATATTTCTGCAGTTTGCCTACTATAACATTGGGAATTATACACAGGC 990  
QY 320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
Db 991 TGT-TGAATGTGCCAAGACCTATCTTCTTCTTCTCCCAATGACGAGGTGATGAACCAA 1049  
QY 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG 360  
Db 1050 ATTTGGCTTATATGACAGCTATGCTTGGAGAGGAACACACACAGATCCATCGGCCCGCTG 1109  
QY 360 luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysLeuLeuPhePheAlaAr 380  
Db 1110 AGAGTGCACCAAGAGTACCAGACGAGACCTACT-GAAAAAGAACTGCTTTTCTTCTGCTT 1168

```
QY 380 YrAspValPheGlyIleProPheValAspProAspSerThrProGluGluValIleP 400
Db 1169 ATGATGTTTTTGAATATCCCTTTGTGATCCGATTCATGAGTCTCCAGAGAGTGAATC 1228
QY 400 rOlysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420
Db 1229 CCAAGAGATTGCAAGAGAAACAGAGTTCAGAACCGGAAACAGCGGTACGCATCTCCAGG 1288
QY 420 lulleGlyAsnLeuMetLysGlnLysGlnLysThrValGluGluLysThrLysGluSerL 440
Db 1289 AGATTGGGAACCTTATGAAGGAATCGAGACCTTGTGGAGAGAGAACCAAGAGTCAAC 1348
QY 440 euAspValSerArgLeuThrArgGlnGlyGlyProLeuLeuTyrGluGluIleSerLeu 460
Db 1349 TGGATGTGACAGACTGACCCGGAGGTGGCCCTGCTGTATGAAGCATCAGTCTCA 1408
QY 460 hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer 479
Db 1409 CCATGAATCTCAAACTCCTGAATGGTTC-CCAGCGGGTGGTGTGATGACCGCGTAATCTCT 1467
QY 480 AspHisGlnCysGlnGlnLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499
Db 1468 GACCACAGTGTGAGGAGCTGCAGAGACTGACCAATGTGSCAGCAACCTCAGGAGATGGC 1527
QY 500 TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe 519
Db 1528 TACCGGGGTTCAGACCTCCCCACATACCTCCCAATGAAAGTTCTATGGTGTCACTGTCTTC 1587
QY 520 LysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyr 539
Db 1588 AAAGCCCTCAAGCTGGGGCAAGGAAAGTTCCTCTGAGAGTGGCCACCTGTACTAC 1647
QY 540 AsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeu 559
Db 1648 AACGTGACGAGAGAGTGGCGGCATCATGAGTCTCTTCCGCTGGATAGCCCTC 1707
QY 560 TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArg 579
Db 1708 TACTTTTCTCTACTCTCATCTGTTGCGGCACTGCCATCGCAAGAGTTCAGGAGAGG 1767
QY 580 LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599
Db 1768 AAGGATGATGTATCCAGTCCAGTGGACAATGCAATCTGATGCCGAGACCTCGTG 1827
QY 600 CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619
Db 1828 TGTGTCAAAGAGCCCCAGAGCTACACCTTCGGGACTACAGCGCCATCCCTTACCTAAAT 1887
QY 620 GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla 639
Db 1888 GGGGACTTCGATGGCGGAAACTTTTATTCTACTGAATGGATGCCAAGACCGTGACGGCA 1947
QY 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659
Db 1948 GAGTGTGAGCCTCAGTGTGGAAGAGCGTGGATTCTCTTCAGGCACTGAAACCCACAT 2007
QY 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuAsp 679
Db 2008 GGAGTGAAGGTGTTCACAGGGGCGAGCGTGTGCCATCCCTGTGTTCACCCCTGGAC 2067
QY 680 ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer 699
Db 2068 CCTGCACACAGCGGGGACAGGGTGCAGGCAGATGACCTGGTGAAGATGCTCTTCAGC 2127
QY 700 ProGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGlu 719
Db 2128 CCAGAGAGATGGACCTCTCCAGAGAGCAGCCCTGGATGCCAGCAGGGGCCCCCGGAA 2187
QY 720 ProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736
Db 2188 CCTGCACAGAGTCTCTCTCAGGAGTGAAATCGAAGCCCAAGGATGAGCTA 2238
```

## RESULT 2

```
BUI77586
LOCUS BUI77586
DEFINITION AGENCOURT_7954111 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6109712
5', mRNA sequence.
ACCESSION BUI77586
VERSION BUI77586.1 GI:22691570
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 848)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2956 row: a column: 09
High quality sequence stop: 635.
FEATURES
Location/Qualifiers
1..848
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6109712"
/lab_host="DH10B (phage-resistant)"
/tissue_type="melanotic melanoma, cell line"
/clone_lib="NIH MGC 112"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
```

## ORIGIN

```
Alignment Scores:
Pred. No.: 8,61e-221 Length: 848
Score: 245.00 Matches: 245
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 33.29% Indels: 0
DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x BUI77586 (1-848)
QY 471 ArgValValMetAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThr 490
Db 2 CCGGGTGGTATGAGCGCGTAACTCTGACACAGAGTGTCCAGAGCTGCAGAGATGACC 61
QY 491 AsnValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsn 510
Db 62 AATGTGGCAGCAACCTCAGGAGATGGTACCGGGGTCCAGACCTCCCACTACTCCCAAT 121
QY 511 GluLysPheTyrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGluLysVal 530
Db 122 GAAAGATTCTATGGTGTCTACTGTCTTCAAGCCCTCAAGCTGGGCAAGAGCAAGTT 181
QY 531 ProLeuGlnSerAlaHisLeuTyrTyrAsnValThrGluLysValArgArgIleMetGlu 550
Db 182 CCTCTGCAGATGCCACCTGTACTACAGTGCAGGAGAGAGTGGCGGCATCATGGAG 241
QY 551 SerTyrPheArgLeuAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThr 570
```

Db 242 TCCTACTTCCCGCTGGATACGCCCTCTACTTTTCTACTCTCATCTGGTGGCGCACT 301  
 Qy 571 AlalleGluValGlnAlaGluArgLysAspSerHisProValHisValAspAsn 590  
 Db 302 GCCATCGAAGAGGTCAGGACAGAGAGAGGATGATGATCATCCAGTCCAGTGGACAAC 361  
 Qy 591 CysIleLeuAsnAlaGluThrLeuValCysValLysGluProProAlaTyrThrPheArg 610  
 Db 362 TGCATCTCTGAATGCCGAGACCTCGTGTGTGTCAAGAGGCCCGCAGCTACACTTCCGC 421  
 Qy 611 AspTyrSerAlaIleLeuTyrLeuAsnGlyValPheAspGlyGlyAsnPheTyrPheThr 630  
 Db 422 GACTACAGCGCATCTTACTTAATGGGACTTCGATGGGGAACCTTTTATTTCACT 481  
 Qy 631 GluLeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGly 650  
 Db 482 GAACTGGATGCAAGACCGTGACGGCAGAGGTGACGCTCAGTGTGGAAGAGCGGTGGGA 541  
 Qy 651 PheSerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCys 670  
 Db 542 TTCTCTCAGGCACTGAAACCCACATGGAGTGAAGCTGTTCACCGGGGCGAGCGCTGT 601  
 Qy 671 AlalleAlaLeuTyrPheThrLeuAspProArgHisSerGluArgAspArgValGlnAla 690  
 Db 602 GCCATCGCCCTGTGTTACCTCGACCTCGACACAGCGAGCGGGACAGGGTGCAGGCA 661  
 Qy 691 AspAspLeuValLysMetLeuPheSerProGluGluMetAspLeuSerGlnGluPro 710  
 Db 662 GATGACCTGGTGAAGATGCTCTTCAGCCCAAGAGATGAGCTCTCCCGAGGAGCGCC 721  
 Qy 711 LeuAspAlaGlnGln 715  
 Db 722 CTGGATGCCCGAGCAG 736  
 RESULT 3  
 BUI47415  
 LOCUS  
 DEFINITION AGENCOURT 8071055 NIH\_MGC\_112 Homo sapiens cDNA clone IMAGE:6090065  
 5', mRNA sequence.  
 ACCESSION BUI47415  
 VERSION BUI47415.1 GI:22660947  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NIH-MGC <http://mgi.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-rc@mail.nih.gov](mailto:cgabbs-rc@mail.nih.gov)  
 Tissue Procurement: DCTD/DTF  
 CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LLCM2330 row: n column: 18  
 High quality sequence stop: 635.  
 Location/Qualifiers  
 1. .905  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6090065"  
 /tissue\_type="melanotic melanoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_112"  
 /notes="Organ: skin; Vector: pOTB7; Site: 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned

into EcoRI/XhoI sites using the following 5' adaptor:  
 GCACGAG(G). Library constructed by Ling Hong in the  
 laboratory of Gerald M. Rubin (University of California,  
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
 Superscript II RT (Life Technologies). Note: this is a  
 NIH\_MGC Library."

## ORIGIN

Alignment Scores:  
 Pred. No.: 8,186-220 Length: 905  
 Score: 244.00 Matches: 244  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 33.15% Indels: 0  
 DB: 13 Gaps: 0  
 US-10-045-815-4 (1-736) x BUI47415 (1-905)  
 Qy 471 AsgValValMetAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThr 490  
 Db 2 CGGGTGGTGTGGAGCGGGTAAATCTCTGACCAAGTGTGAGGCTGCAGAGACTGACC 61  
 Qy 491 AsnValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsn 510  
 Db 62 AATGTGGCAGCAACCTCAGGAGATGGCTACCGGGGTGAGACTCCCCACATATCCCAAT 121  
 Qy 511 GluLysPheTyrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGluLysVal 530  
 Db 122 GAAAGTTTCTATGGTGTCTCTCTTCAAGCCCTCAAGCTGGGGCAGAAGGCAAGTT 181  
 Qy 531 ProLeuGlnSerAlaHisLeuTyrTyrAsnValThrGluLysValaArgArgileMetGlu 550  
 Db 182 CCTCTGCAGAGTGCACCTGTACTTACAACTGACGAGAGAGGTGCGGCGCATCATGGAG 241  
 Qy 551 SerTyrPheArgLeuAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThr 570  
 Db 242 TCCTACTTCCGCTGGATACGCCCTCTACTTTTCTACTCTCATCTGTGTGGCGCACT 301  
 Qy 571 AlalleGluValGlnAlaGluArgLysAspSerHisProValHisValAspAsn 590  
 Db 302 GCCATCGAAGAGTCCAGCGCAGAGAGGATGATGATCATCCAGTCCACGTGGACAAC 361  
 Qy 591 CysIleLeuAsnAlaGluThrLeuValCysValLysGluProProAlaTyrThrPheArg 610  
 Db 362 TGCATCTCTGAATGCCGAGACCTCGTGTGTGTCAAGAGGCCCGCAGCTACACTTCCGC 421  
 Qy 611 AspTyrSerAlaIleLeuTyrLeuAsnGlyValPheAspGlyGlyAsnPheTyrPheThr 630  
 Db 422 GACTACAGCGCATCTTACTTAATGGGACTTCGATGGGGAACCTTTTATTTCACT 481  
 Qy 631 GluLeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGly 650  
 Db 482 GAACTGGATGCAAGACCGTGACGGCAGAGGTGACGCTCAGTGTGGAAGAGCGGTGGGA 541  
 Qy 651 PheSerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCys 670  
 Db 542 TTCTCTCAGGCACTGAAACCCACATGGAGTGAAGCTGTTCACCGGGGCGAGCGGTGT 601  
 Qy 671 AlalleAlaLeuTyrPheThrLeuAspProArgHisSerGluArgAspArgValGlnAla 690  
 Db 602 GCCATCGCCCTGTGTTACCTCGACCTCGACACAGCGAGCGGGACAGGGTGCAGGCA 661  
 Qy 691 AspAspLeuValLysMetLeuPheSerProGluGluMetAspLeuSerGlnGluPro 710  
 Db 662 GATGACCTGGTGAAGATGCTCTTCAGCCCAAGAGATGAGCTCTCCCGAGGAGCGCC 721  
 Qy 711 LeuAspAlaGln 714  
 Db 722 CTGGATGCCCGAG 733

## RESULT 4

BQ677232

LOCUS

929 bp mRNA linear EST 15-JUL-2002



```

DEFINITION AGENCOURT_8494737 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6294988
5', mRNA sequence.
ACCESSION B0677232
VERSION B0677232.1 GI:21789911
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 929)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/BTP
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM2500 row: i column: 05
High quality sequence start: 5
High quality sequence stop: 638.
FEATURES
    source
        1..929
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:6294988"
            /tissue_type="melanotic melanoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH_MGC_112"
            /notes="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
            EcoRI; cDNA made by oligo-dT priming. Directionally cloned
            into EcoRI/XhoI sites using the following 5' adaptor:
            GGACAG(G). Library constructed by Ling Hong in the
            laboratory of Gerald M. Rubin (University of California,
            Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
            Superscript II RT (Life Technologies). Note: this is a
            NIH_MGC Library."
ORIGIN
Alignment Scores:
Pred. No.: 5,12e-216 Length: 929
Score: 240.00 Matches: 240
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 32.61% Indels: 0
DB: 13 Gaps: 0
US-10-045-815-4 (1-736) x B0677232 (1-929)
QY 497 GlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyVal 516
Db 17 GGAGATGGCTACCGGGGTGAGACCTCCCACTACTCCCAATGAAAGTTCATGTGTGC 76
QY 517 ThrValPheLysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHis 536
Db 77 ACTGTCTTCAAGCCCTCAAGCTGGGGCAAGAGGCAAGATTCCTCTGAGAGTGGCCAC 136
QY 537 LeuTyrTyrAsnValThrGluLysValArgArgileMetGluSerTyrPheArgLeuAsp 556
Db 137 CTGTACTACACGTGACGGAGAGGTGGCGCATCATGAGTCTTCCGCTGGAT 196
QY 557 ThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGlnValGln 576
Db 197 ACGCCCCCTTACTTTTCTTACTCTCATCTGCTGGTGGCGCACTGCCATCGAAGAGGTCAG 256
QY 577 AlaGluArgLysAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGlu 596
Db 257 GCAGAGGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 316

```

```

QY 597 ThrLeuValCysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeu 616
Db 317 ACCCTCGTGTGTCTCAAGAGAGCCCGACGCTACACCTTCCGGAGCTAGAGGCCATCTT 376
QY 617 TyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThr 636
Db 377 TACCTAAATGGGAGCTTCGATGGCGGAAACTTTTATTTCACTGAATCGATGCCAAGACC 436
QY 637 ValThrAlaGluValGlnProGlnCysGlyValArgAlaValGlyPheSerSerGlyThrGlu 656
Db 437 GTGACGGCAGAGGTGCACGCTCAGTGTGGAAGAGCCGTTGGGATTTCTTTCAGGCTGAA 496
QY 657 AsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPhe 676
Db 497 AACCCACATGGAGTGAAGGCTGTACACAGGGGGCAGCGCTGTGCCATCGCCTGTGGTTC 556
QY 677 ThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMet 696
Db 557 ACCCTGGACCTTCGACACAGCGAGCGGAGGTGCGAGGAGATGACCTGTGTGAAGATG 616
QY 697 LeuPheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGly 716
Db 617 CTCCTCAGCCAGAGAGATGGACCTCTCCAGGAGCAGCCCTGGATGCCAGCAGGCG 676
QY 717 ProProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736
Db 677 CCCCCGAACTCGACAAAGAGTCTCTCTCAGGCACTGAATCGAAGCCCAAGATGAGCTA 736

```

```

RESULT 5
BUI79488
LOCUS BUI79488
DEFINITION BUI79488 899 bp mRNA linear EST 04-SEP-2002
5', mRNA sequence.
ACCESSION BUI79488
VERSION BUI79488.1 GI:22693472
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 899)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/BTP
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM2291 row: i column: 24
High quality sequence stop: 627.
FEATURES
    Location/Qualifiers
        1..899
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:6074975"
            /tissue_type="melanotic melanoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH_MGC_112"
            /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
            EcoRI; cDNA made by oligo-dT priming. Directionally cloned
            into EcoRI/XhoI sites using the following 5' adaptor:
            GGACAG(G). Library constructed by Ling Hong in the
            laboratory of Gerald M. Rubin (University of California,
            Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
            Superscript II RT (Life Technologies). Note: this is a
            NIH_MGC Library."

```

ORIGIN		NIH-MGC http://mgc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue procurement: DCTD/DTP cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLCW2399 row: f column: 23 High quality sequence stop: 588. location/Qualifiers	
FEATURES		1..884 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:6252694" /tissue_type="melanotic melanoma, cell line" /lab_host="DHI08 (phage-resistant)" /clone_lib="NIH_MGC_112" /note="Organ: Skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."	
ORIGIN		Alignment Scores: Pred. No.: 1.4e-206 Length: 884 Score: 230.00 Matches: 230 Percent Similarity: 100.00% Conservative: 0 Best Local Similarity: 100.00% Mismatches: 0 Query Match: 31.25% Indels: 0 DB: 13 Gaps: 0 US-10-045-815-4 (1-736) x BUI190877 (1-884)	
QY	498	AspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThr	517
DB	2	GATGGCTACCGGGTCAGACCTCCCAATGAAAAAGTTCTATGGTGTCACT	61
QY	518	ValPheLysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeu	537
DB	62	GTCTTCAAGGCCCTCAAGCTGGGGCAAGAGCAAGTTCTCTGACAGATGCCCATCTG	121
QY	538	TyTyrAsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThr	557
DB	122	TACTACACGTGACGAGAGAGTGGCGGCATCATCGAGTCTCTACTTCCGCTGGATACG	181
QY	558	ProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAla	577
DB	182	COCCTCTACTTTTTCCTACTCTCATCTGGTGTGCCGCACCTCCATCGAAGAGTCCAGCA	241
QY	578	GluArgLysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThr	597
DB	242	GAGAGAGAGGATGATGATCATCCAGTCCAGTGGACAACTGCATCTGATGCCGAGACC	301
QY	598	LeuValCysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyr	617
DB	302	CTCGTGTGTGTCAAGAGCCCGCCAGCCTACACCTTCCGCGACTACAGCGCATCTTTAC	361
QY	618	LeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrVal	637
DB	362	CTAAATGGGACTTCGATGGCGGAAACTTTATTTTCACTGACTGGATGCCAGACCGTG	421
QY	638	ThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsn	657

AUTHORS  
TITLE  
JOURNAL  
COMMENT

NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: DCTD/DPV  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed By: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing By: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov

Plate: LLCW2399 row: f column: 23  
High quality sequence stop: 588.

FEATURES  
source

1..884  
location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6252694"  
/tissue\_type="melanotic melanoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_112"  
/note="Organ: skin; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.: 1.4e-205 Length: 884  
Score: 230.00 Matches: 230  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 31.25% Indels: 0  
DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x BUI190877 (1-884)

QY 498 AspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThr 517  
DB 2 GATGGCTACCGGGTCAGACCTCCCCACATATCTCCCAATGAAAGTTCTATGGTCTCACT 61  
QY 518 ValPheLysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeu 537  
DB 62 GTCTTCAAGCCCTCAAGCTGGGCAAGAGCAAGTTCTCTGCAGAGTCCACCTG 121  
QY 538 TyrTyrAsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThr 557  
DB 122 TACTACAACTGACGAGAGAGGTGGCGCATCATGGAGTCTCTTCCGCTCGATACG 181  
QY 558 ProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAla 577  
DB 182 CCCCTCTACTTTTCTTCTACTCTCATCTGTGTGCGCACTGCCATCGAAGAGTCCAGGCA 241  
QY 578 GluArgLysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThr 597  
DB 242 GAGAGGAGGATGATAGTATCCAGTCCAGCTGGACAACTCATCTCTGATGCCGAGACC 301  
QY 598 LeuValCysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyr 617  
DB 302 CTCGTGTGTGTCAAAGAGCCCCCAGCCCTACACCTTCCCGCACTACAGCGCCATCTTTTAC 361  
QY 618 LeuAsnGlyAspPheAspGlyGlyValAsnPheTyrPheThrGluLeuAspAlaLysThrVal 637  
DB 362 CTAATGGGACTTCGATGGCGGAACTTTTATTTCTGTAACCTGGATGCCAAGACCGTG 421  
QY 638 ThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsn 657

422 ACAGCAGAGTGCAGCCTCAGTGTGGAAGAGCGGTGGATTCTCTTCAGGCACCTGAAAC 481  
 658 ProHisGlyValIysAlaValThrArgGlyGlnAtcCysAlaIleAlaLeuTrpPheThr 677  
 482 CCATGTGAGTGAAGCTGTACACAGGGGAGCGGTGTGCCATGCCCTGTGGTTACCC 541  
 678 LeuAspProArgHisSerGluArgAspA-gValGlnAlaAspAspLeuValIysMetLeu 697  
 542 CTGACCTTCACACAGCAGCGGGACAGGGTGCAGGCAGATGACCTGTGTGAAGATGCTC 601  
 698 PheSerProGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyPro 717  
 602 TTGAGCCCAAGAGATGACCTCTCCAGGAGCAGCCCTGGATGCCAGCAGGGGCC 661  
 718 ProGluProAlaGlnGluSerLeuSerGly 727  
 662 CCCGAACCTGCACAGAGTCTCTCTCAGGG 691

BF206567 937 bp mRNA linear EST 06-NOV-2000  
 601870908F1 NIH\_MGC\_19 Homo sapiens cDNA clone IMAGE:4101721 5',  
 mRNA sequence.

ACCESSION BF206567.1 GI:11100153  
 VERSION EST.  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 937)  
 NIH-MGC http://mgi.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LLC9373 row: 0 column: 02  
 High quality sequence stop: 700.  
 Location/Qualifiers  
 1..937  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4101721"  
 /tissue\_types="neuroblastoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_19"  
 /notes="Organ: brain; Vector: pOTB7; Site:1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5',  
 adaptor: GGACGAG(G). Library constructed by Ling Hong  
 in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."

ORIGIN

Alignment Scores:  
 Pred. No.: 1.7e-172 Length: 937  
 Score: 194.00 Matches: 194  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 26.36% Indels: 0  
 DB: 10 Gaps: 0

US-10-045-815-4 (1-736) x BF206567 (1-937)

492 ValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGlu 511  
 4 GTGGCAGCAACTCAGGAGATGGCTACCGGGGTCCAGACCTCCCCACATACCTCCCAATGAA 63  
 512 LysPheTyrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGluGlyLysValPro 531  
 64 AAGTTCTATGTGTGTCACCTGTTTCAAGCCCTCAAGCTGGGGCAAGAGCAAGATTCCT 123  
 532 LeuGlnSerAlaHisLeuTyrTyrAsnValThrGluLysValArgArgIleMetGluSer 551  
 124 CTGAGAGTGCACCACTGTACTACAGTGCAGGAGAGAGTGCAGGCATCATCGAGTCC 183  
 552 TyrPheArgLeuAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAla 571  
 184 TACTTCGGCTGGATAGCCCTCTACTTTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 243  
 572 IleGluGluValGlnAlaGluArgLysAspSerHisProValHisValAspAsnCys 591  
 244 ATCGAAGAGGTCCAGGAGAGAGAGATATAGTCATCCAGTCCAGCTGGACACTGC 303  
 592 IleLeuAsnAlaGluThrLeuValCysValLysGluProProAlaTyrThrPheArgAsp 611  
 304 ATCTGAATGCCGAGACCTCGTGTGTCAAGAGCCCCCAGCTACACCTTCGGGAC 363  
 612 TyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGlu 631  
 364 TACAGCGCCATCTTACTTAATGGGACTTCGATGGCGAAACTTTTATTTCCTGAA 423  
 632 LeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPhe 651  
 424 CTGATGCAAGACCGTGACGGCAGAGGTGAGCTCAGTGTGGAAGAGCGGTGGATTC 483  
 652 SerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAla 671  
 484 TCTTCAGGCACGTAAACCCACATGGAGTGAAGCTGTCAACAGGGGGGAGCGGTGTGCC 543  
 672 IleAlaLeuTrpPheThrLeuAspProArgHisSerGluArg 685  
 544 ATGCGCCCTGTGGTTCACCTCGACCCCTGCACACAGCGAGCGG 585

BO678884 989 bp mRNA linear EST 15-JUL-2002  
 AGENCOURT\_8497890 NIH\_MGC\_112 Homo sapiens cDNA clone IMAGE:6257808  
 5', mRNA sequence.

ACCESSION BO678884  
 VERSION BO678884.1 GI:21791563  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 989)  
 NIH-MGC http://mgi.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-remail.nih.gov  
 Tissue Procurement: DCTD/DTF  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLCM2412 row: 1 column: 01  
 High quality sequence start: 46  
 High quality sequence stop: 404.  
 Location/Qualifiers  
 1..989  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"

/clone="IMAGE:6257808"  
 /tissue\_type="melanotic melanoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_112"  
 /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

## ORIGIN

Alignment Scores:  
 Pred. No.: 1.41e-170 Length: 989  
 Score: 192.00 Matches: 192  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 26.09% Indels: 0  
 DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x B0678884 (1-989)

Qy 470 GlnArgValValMetAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeu 489  
 Db 12 CAGCGGTGGTGGATGGAGCGCGTAACTCTGACCAGAGTGTGAGGAGTGCAGAGACTG 71  
 Qy 490 ThrAsnValAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrPro 509  
 Db 72 ACCAATGTGGCAGCAACTCAGGAGATGGTACCGGGGTGAGACCTCCCGACATATCTCC 131  
 Qy 510 AsnGluLysPheTyrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGluGlyLys 529  
 Db 132 AATGAAAGTTCTATGTGTGTCACCTGCTTCAAGCCCTCAAGTGGGGCAAGAGGCANA 191  
 Qy 530 ValProLeuGlnSerAlaHisLeuTyrAsnValThrGluLysValArgArgIleMet 549  
 Db 192 GTTCTCTGCAGAGTGCACCTGTACTCAACGTGACGAGAGGTGGCGGCATCATG 251  
 Qy 550 GluSerTyrPheArgLeuAspThrProLeuTyrPheSerTyrSerHisLeuValCysArg 569  
 Db 252 GAGTCTCTCTCCGCTGGATACGCCCTCTACTTTCTACTCTCTCTCTCTCTCTCTCTCT 311  
 Qy 570 ThrAlaIleGluGluValGlnAlaGluArgLysAspAspSerHisProValHisValAsp 589  
 Db 312 ACTGCCATCGAAGAGTCCAGGCAGAGAGGATGATGATCATCATCCAGTCCACGTGGAC 371  
 Qy 590 AsnCysIleLeuAsnAlaGluThrLeuValCysValLysGluProProAlaTyrThrPhe 609  
 Db 372 AACTGTCATCTTGATGCGGACACCTCGTGTGTCTAAGAGCCCGCCAGCCTACACCTTC 431  
 Qy 610 ArgAspTyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPhe 629  
 Db 432 CGCGACTACAGCGCCATCCCTTACCTAAATGGGACTTCGATGGCGAAACTTTTATTC 491  
 Qy 630 ThrGluLeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaVal 649  
 Db 492 ACTGAATGTATGCAAGACCGTGACGGCAGAGGTGCAGCCTCAGTGTGGAAGCCGTG 551  
 Qy 650 GlyPheSerSerGlyThrGluAsnProHisGlyVal 661  
 Db 552 GGATTCCTCTCAGGACTGAAACCCACATGGAGTG 587

## RESULT 9

BE294207  
 LOCUS 705 bp mRNA linear EST 20-JUL-2000  
 DEFINITION 601172737F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:3528273 5', mRNA sequence.  
 ACCESSION BE294207  
 VERSION BE294207.1 GI:9177841  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euteria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 705)

## AUTHORS

NIH-MGC http://mgc.ncbi.nlm.nih.gov/.

## TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

## JOURNAL

Unpublished (1999)

## COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLCW197 row: i column: 10

High quality sequence stop: 641.

Location/Qualifiers

1..705

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:3528273"

/tissue\_type="rhodomyosarcoma"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_17"

/note="Organ: muscle; Vector: pOTB7; Site 1: EcoRI;

Site 2: XhoI; cDNA made by oligo-dt priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGACGAG(G). Size-selected >500bp

for average insert size 1.8kb. Library constructed by

Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

## ORIGIN

Alignment Scores:  
 Pred. No.: 7.56e-169 Length: 705  
 Score: 190.00 Matches: 203  
 Percent Similarity: 99.51% Conservative: 0  
 Best Local Similarity: 99.51% Mismatches: 0  
 Query Match: 25.82% Indels: 1  
 DB: 10 Gaps: 0

US-10-045-815-4 (1-736) x BE294207 (1-705)

Qy 486 LeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSer 505  
 Db 2 CTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGCTACCGGGTCCAGACCTCC 61  
 Qy 506 ProHisThrProAsnGluLysPheTyrGlyValThrValPheLysAlaLeuLysLeuGly 525  
 Db 62 CCACATATCTCCCAATGAAAGTTCTATGGTGTCTACTGTCTTCAAGGCCCTCAAGCTGGG 121  
 Qy 526 GlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsnValThrGluLysVal 545  
 Db 122 CAGNAGGCAAGTTCCTCTGCAGAGTGCACCTGCTACTACACGTCGACGAGAGGTG 181  
 Qy 546 ArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyrPheSerTyrSerHis 565  
 Db 182 CGGCGCATCATGAGTCTCTTCCGCTGTGATACGCCCTCTACTTTCCTACTCTCAT 241  
 Qy 566 LeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLysAspAspSerHisPro 585  
 Db 242 CTGTGTGCCGCACTGCCATCCAGAGGTCCAGCAGAGAGAGAGATGATGATCATCCA 301  
 Qy 586 ValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCysValLysGluProPro 605  
 Db 302 GTCCACGTGGCAACTGTCATCTGAATGCCGAGACCTCGTGTGTGTCAAGAGACCCCCA 361  
 Qy 606 AlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGlyGly 625  
 Db 362 GCCTACACCTTCCCGGACTACAGGCCCATCTTTTACCTAAATGGGACTTCGATGGCGGA 421

626 AsnPhetYrPheThrGluLeuAspAlaLysThrValThrAlaGluValGlnProGlnCys 645  
422 AACCTTTTATTTCACCTGAACCTGGATGTCACACCGTGACCGGACGAGCTCAGTGT 481  
646 GlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGlyValLysAlaValThr 665  
482 GGAAGAGCGCTGGGATTCTCTTCAGGCACCTGAAACCCACATGAGTGAAGGCTGTACCC 541  
666 ArgGlyGlnArgCysAlaIleAlaLeu-TrpPheThrLeuAspProArgHisSerGluAr 685  
542 AGGGGGCAGCGCTGTGCCATCGCCCTGTGTTTCACCTGACCTCGACACAGCGAGCG 601  
685 gAspArgVal 688  
602 GCACAGGGTG 611

RESULT 10  
BQ680853 885 bp mRNA linear EST 15-JUL-2002  
LOCUS AGENCOURT\_8192213 NIH\_MGC\_112 Homo sapiens cDNA clone IMAGE:6260027  
DEFINITION 5', mRNA sequence.  
ACCESSION BQ680853  
VERSION BQ680853.1 GI:21793532  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 885)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: DCTD/DTP  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM2418 row: h column: 12  
High quality sequence stop: 534.

FEATURES  
source  
1..885  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6260027"  
/tissue types="melanotic melanoma, cell line"  
/lab host="DH10B (phage-resistant)"  
/clone lib="NIH\_MGC\_112"  
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGCAG(G). Library constructed by Ling Hong in the  
laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH\_MGC Library."

ORIGIN  
Alignment Scores:  
Pred. No.: 6.71e-166 Length: 885  
Score: 187.00 Matches: 187  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 25.41% Indels: 0  
DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x BQ680853 (1-885)

471 ArgValValMetAspGlyValIleSerAspHisGluCysGlnGluLeuGluArgLeuThr 490  
2 CGGGTGGTGATGGACGGGGTAATCTCTGACCACGAGTGTCCAGGAGCTCAGAGACTGACC 61  
491 AsnValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsn 510  
62 AATGTGGCAGCAACCTCAGGAGATGGCTACCGGGGTCCAGACTCCCCACATACTCCCAAT 121  
511 GluLysPheTyrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGluGlyLysVal 530  
122 GAAAAGTTCTATGGTGTCTCTTCAAGCCCTCAAGCTGGGGCAAGGCAAGTT 181  
531 ProLeuGlnSerAlaHisLeuTyrTyrAsnValThrGluLysValArgArgIleMetGlu 550  
182 CCTCTGCAGAGTCCCACTGTACTACCACTGACGAGAGAGGTCGGCGCATCATGGAG 241  
551 SerTyrPheArgLeuAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThr 570  
242 TCCTACTTCGGCTCGATAGCCCTCTACTTTCTCTACTCTCTCTCTCTCTCTCTCTCT 301  
571 AlaIleGluGluValGluAlaGluArgLysAspSerHisProValHisValAspAsn 590  
302 GCCATCGAAGAGGTCCAGCAGAGAGGAGGATGATGATCCAGTCCAGTGGGACAAAC 361  
591 CysIleLeuAsnAlaGluThrLeuValCysValLysGluProProAlaTyrThrPheArg 610  
362 TGCATCTCTGAATGCCGAGACCTCTGTGTGTCAAGAGAGCCCGACCTTCCGCTCCGC 421  
611 AspTyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThr 630  
422 GACTACAGCGCCATCTTTTACCTAAATGGGACTTCGATGGCGGAACTTTTATTTCAT 481  
631 GluLeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGly 650  
482 GAACTGGATGCCAGACCTGTGACGACAGGTCGAGCTCAGTGTGGAGAGCCCTGGGA 541  
651 PheSerSerGlyThrGluAsn 657  
542 TTCCTCTTCAGGCACTGAAAC 562

RESULT 11  
BQ676738 870 bp mRNA linear EST 15-JUL-2002  
LOCUS AGENCOURT\_8196745 NIH\_MGC\_112 Homo sapiens cDNA clone IMAGE:6263574  
DEFINITION 5', mRNA sequence.  
ACCESSION BQ676738  
VERSION BQ676738.1 GI:21789417  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 870)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: DCTD/DTP  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM2427 row: 1 column: 07  
High quality sequence stop: 689.

FEATURES  
source  
1..870  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6263574"

```

/tissue type="melanotic melanoma, cell line"
/lab host="DH10B (phage-resistant)"
/clone.lib="NIH_MGC_112"
/notes="Organ: skin; Vector: pOTB7; Site: 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 5.13e-164 Length: 870
Score: 185.00 Matches: 255
Percent Similarity: 98.84% Conservative: 0
Best Local Similarity: 98.84% Mismatches: 1
Query Match: 25.14% Indels: 3
DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x B0676738 (1-870)
QY 412 GluThrAlaValArgIleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThrLeu 431
Db 1 GAAACAGCGGACCGCATCTCCAGAGATGGGAACCTTATGAAGGAATCGAGACCCIT 60
QY 432 ValGluGluThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyGlyPro 451
Db 61 GTGAAGAGAGACCAAGAGTCACTGGATGTGAGCAGACTGACCCGGAGGTGGCCCC 120
QY 452 LeuLeuThrGluGlyIleSerLeuThrMetAsnSerLysLeuAsnGlyTyr-GlnAr 471
Db 121 CTGCTGTATGAAGCATCAGTCTCACCATGAACCTCAAACTCTCGAATGTTTC-CCAGCG 179
QY 471 qValValMetAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAs 491
Db 180 GTGTGTATGAGCGCGGTATCTCTGACCAGAGTGTACAGAGCTGACAGAGTGCACCA 239
QY 491 nValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnG 511
Db 240 TGTGGCAGCAACCTCAGAGATGCTACCGGGGTACAGCCTCCCCACATCTCCCAATGA 299
QY 511 uLysPheThrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGluGlyValPr 531
Db 300 AAAGTTCTATGTGTCTACTGCTTCAAGGCCCTCAAGCTGGGGCAAGAGCGCAAGTTC 359
QY 531 oLeuGlnSerAlaHisLeuThrValThrValThrGluLysValArgIleMetGluSe 551
Db 360 TCTCAGAGTGCACCATCTGTACTACAACGTGACGAGAGGTGGCGGCATCATGGAGTC 419
QY 551 rTyrPheArgLeuAspThrProLeuThrPheSerTyrSerHisLeuValCysArgThrAl 571
Db 420 CTACTTCGCGCTGGATACGCCCTCTACTTTCTCTACTCTCATCTGTGTGGCGACTGC 479
QY 571 aileGluGluValGlnAlaGluArgLysAspSerHisProValHisValAspAsnCy 591
Db 480 CATCGAAGAGGTCCAGGACAGAGGAAGGATGATAGTCACTCCAGTCCAGCTGGACAACTG 539
QY 591 sileLeuAsnAlaGluThrLeuValCysValLysGluProProAlaTyrThrPheArgAs 611
Db 540 CATCCTGAATGCCAGACCCCTCGTGTGTGTAAGAGAGCCCGCCAGCTACACCTTCGCGCA 599
QY 611 pTyrSerAlaIleLeuThrLeuAsnGlyAspPheAspGlyGlyAsnPheThrG 631
Db 600 CTACAGCGCCATCTTTACCTAATGGGACTTCGATGGCGGAACCTTTATTCTACTGA 659
QY 631 uLeuAspAlaLysThrValThrAlaGluValGlnPro-GlnCysGlyArgAlaValGlyP 651
Db 660 ACTGGATGCCAAGACCGGTGACGGCAGAGGTGACGCCNTCAGTGTGAAGAGCGCGTGGAT 719
QY 651 heSerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGly 667

```

```

Db 720 TCTCTTCAGGCACCTGAAACCCACCATGGAGTGAAGGCTGTCCACGAGGGG 769

RESULT 12
BUI48566
LOCUS
DEFINITION
BUI48566 899 bp mRNA linear EST 03-SEP-2002
AGENCOURT.9675673 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6380366
5', mRNA sequence.
BUI48566
VERSION
BUI48566.1 GI:22662098
EST.
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2568 row: n column: 15
High quality sequence stop: 724.
Location/Qualifiers
1..899
/organism="Homo sapiens"
/db_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6380366"
/tissue type="carcinoma, cell line"
/lab host="DH10B (phage-resistant)"
/clone.lib="NIH_MGC_40"
/notes="Organ: prostate; Vector: pOTB7; Site: 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGCAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 3.24e-160 Length: 899
Score: 181.00 Matches: 281
Percent Similarity: 99.29% Conservative: 0
Best Local Similarity: 99.29% Mismatches: 1
Query Match: 24.59% Indels: 2
DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x BUI48566 (1-899)
QY 143 ArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGluLysAla 162
Db 3 CGAGCCCTTCAACTACTCTCAGTCCCTTCAAGATCAACAGTTGGAGAAAGCT 62
QY 163 ValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGlnAsn 182
Db 63 GTTGTGTCAGCACACACCTTCTTGTGGCAATCCTGAGCACATGGAAATGCAGCAAC 122
QY 183 LeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeuGluThr 202
Db 123 CTAGACTATTACCAACCATCTGTGGAGTGAAGAGCGCCGACTTCAAGGATCTTGAGACT 182
QY 203 GlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGlnProGln 222
Db 183 CAACCCCATATGCAAGAAATTTTCGACTGGAGTGGCGACTCTACTCTAGAGAACACCCACAG 242

```

US-10-045-815-4 (1-736) x BU543499 (1-881)

## FEATURES

Published (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-r@mail.nih.gov  
 Tissue Procurement: DCTD/DTP  
 CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
 Plate: LLCM2772 row: m column: 12  
 High quality sequence stop: 735.  
 Location/Qualifiers

## FEATURES

```

Db      662 CAAGGAGTACCGACGACGACCTACTGGAAGAAAGAACTGTTTCTTCGCTTATGATG 721
Qy      382 lPheGlyIleProPheValAspProAspSerTriPThrProGluValIleProLysAr 402
Db      722 TTTTGGAAATCCCTTTGTGGATCCGGATTCATGGACTCCAGAAAGAGTATCCCAAGAG 781
Qy      402 GLeuGlnGluYsGlnYsSerGluAArgGluThrAlaValArgIleSerGlnGluIle 421
Db      782 ATTGCAAGAAACAGAAAGTCAAGACGGGAAACAGCCGTCAGCATCTCCAGGAGATT 839

RESULT 14
LOCUS   BQ678346                      955 bp      mRNA      linear      EST 15-JUL-2002
DEFINITION   AGENCOURT_8500887 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6294282
5', mRNA sequence.
ACCESSION   BQ678346
VERSION     BQ678346.1  GI:21791025
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 955)
AUTHORS     NIH-MGC http://mgi.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: sgapbs-remail.nih.gov
            Tissue Procurement: DCTD/DTF
            cDNA Library Preparation: Rubin Laboratory
            DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Place: LNC2498 row: k column: 19
            High quality sequence stop: 566.
FEATURES   source
            1..955
            Location/Qualifiers
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:6294282"
            /tissue_type="melanotic melanoma, cell line"
            /lab_hosts="DH10B (phage-resistant)"
            /clone_lib="NIH MGC 112"
            /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
            EcoRI; cDNA made by oligo-dT priming. Directionally cloned
            into EcoRI/XhoI sites using the following 5' adaptor:
            GGCACGAG(G). Library constructed by Ling Hong in the
            Laboratory of Gerald M. Rubin (University of California,
            Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
            Superscript II RT (Life Technologies). Note: this is a
            NIH_MGC Library."
ORIGIN
Alignment Scores:
Pred. No.:      1.45e-153      Length:      955
Score:          174.00      Matches:      231
Percent Similarity: 99.14%      Conservative: 0
Best Local Similarity: 99.14%      Mismatches: 1
Query Match:     23.64%      Indels:      2
DB:              13          Gaps:         0

US-10-045-815-4 (1-736) x BQ678346 (1-955)

Qy      412 GluThrAlaValArgIleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThrLeu 431
Db      1   GAAACAGCGGTACGCATCTCCAGACAGTGGGAACCTTATGAAGGAATCGAGACCTT 60
Qy      432 ValGluGlnYsThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyPro 451

```

```

Db      61  GTGGAAGAGAACCAAGGAGTCACTGGATGTGACAGACTGACCCGGGAAGGTGGCCCC 120
Qy      452 LeuLeuTyThrGluGlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyr-GlnAr 471
Db      121 CTGCTGTATGAAGGATCAGTCTCACCATGAATCCAAACTCCTGAATGGTTC-CCAGCG 179
Qy      471 GValValMetAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAs 491
Db      180 GTGTGTATGGAGCGGCTAATCTCTGACCAAGAGTGTGAGGAGTGCAGAGACTGACCAA 239
Qy      491 nValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGl 511
Db      240 TGTGGCAGCAACCTCAGGAGATGGCTAGCGGGGTGAGACCTCCCCACATACTCCCAATGA 299
Qy      511 uLysPheTyThrGlyValThrValPheIysAlaLeuLysLeuGlyGlnGlyLysValPr 531
Db      300 AAAGTTCTAATGGTGTACCTCTTCAAGCCCTCAAGCTGGGGCAAGGAGGCAAGTTCC 359
Qy      531 oLeuGlnSerAlaHisLeuTyThrAsnValThrGluLysValArgArgIleMetGluSe 551
Db      360 TCTGCAGAGTGGCCACCTGTACTCAACGTCACGAGAGAGGTGGCGCATCATGGAGTC 419
Qy      551 rTyThrPheArgLeuAspThrProLeuTyThrPheSerTyThrSerHisLeuValCysArgThrAl 571
Db      420 CTACTTCGCGCTGGATACGCCCTCTACTTTTCTTACTCTCTATCTGGTGTGCCGCACTGC 479
Qy      571 aileGluGluValGlnAlaGluAArgLysAspSerHisProValHisValAspAsnCy 591
Db      480 CATCGAAGAGTCCAGGCAGAGAGAGATGATAGTATCATCCAGTCCAGTGGACAACTG 539
Qy      591 sileLeuAsnAlaGluThrLeuValCysValLysGluProProAlaTyThrPheArgAs 611
Db      540 CATCTCTGAATGCCGAGACCTGTGTGTCTCAAGAGAGCCCGCAGCTACACTTCCCGGA 599
Qy      611 pTyThrSerAlaIleLeuTyThrLeuAsnGlyAspPheAspGlyGlyAsnPheTyThrGl 631
Db      600 CTACAGCGCCATCCCTTTACCTTAATGGGACTTCGATGGCGGAAACTTTTATTTCATGTA 659
Qy      631 uLeuAspAlaLysThrValThrAlaGluValGlnPro 643
Db      660 ACTGGATGCCAAGACCGTGACGGCAGAGAGTGCAGCCT 696

RESULT 15
LOCUS   AW964170                      730 bp      mRNA      linear      EST 01-JUN-2000
DEFINITION   EST376243 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW964170
VERSION     AW964170.1  GI:8154006
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 730)
AUTHORS     Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspar, R., Gay, C.,
            Holt, I. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and
            Quackenbush, J.
            Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
            Unpublished (2000)
JOURNAL     Contact: John Quackenbush
COMMENT     The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@igir.org
            Plate: 198
            Seq primer: Reverse.
            Location/Qualifiers
            1..730
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
FEATURES   source

```



/clone\_lib="WAGE resequences, MAGH"  
/note="Vector: pBluescriptSKm"

ORIGIN

Alignment Scores:

Pred. No.:	5,11e-148	Length:	730
Score:	168.00	Matches:	168
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	22.83%	Indels:	0
DB:	10	Gaps:	0

US-10-045-815-4 (1-736) x AW964170 (1-730)

Qy	569	ArgThrAlaIleGluGluValGlnAlaGluArgLysAspSerHisProValHisVal	588
Db	1	CGCACTGCCATCGAGAGGTCCAGGCGAGAGAGAGATGATGATCATCCAGTCCACGTG	60
Qy	589	AspAsnCysIleLeuAsnAlaGluThrLeuValCysValLysGluProAlaIleThr	608
Db	61	GACAACTGCATCTCGATGCCGAGACCCCTCGTGTGTCAAGAGAGCCCCAGGCTACAC	120
Qy	609	PheArgAspTyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyr	628
Db	121	TTCCGCGACTACAGCGCATCTTTTACCTAAATGGGGACTTCGATGGCGGAAACTTTTAT	180
Qy	629	PheThrGluLeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAla	648
Db	181	TTCACTGAACCTGGATGCCAAGACCGTGCAGGCGAGAGGTGCAGCTCAGTGTGGAAGAGCC	240
Qy	649	ValGlyPheSerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGln	668
Db	241	GTGGGATTCCTTCAGGCACTGAAACCCACATGGAGTGAAGGCTGTACCCAGGGGGCAG	300
Qy	669	ArgCysAlaIleAlaLeuTyrPheThrLeuAspProArgHisSerGluArgAspArgVal	688
Db	301	CGCTGTGCCATCGCCCTGTGGTTACCCCTGGACCCCTGCACACAGCGAGCGGGACAGGGTG	360
Qy	689	GlnAlaAspAspLeuValLysMetLeuPheSerProGluGluMetAspLeuSerGlnGlu	708
Db	361	CAGGCAGATGACCTGTGTGAAGATGCTCTTCAGGCCAGAGAGATGGACCTCTCCAGGAG	420
Qy	709	GlnProLeuAspAlaGlnGlnGlyProGluProAlaGlnGluSerLeuSerGlySer	728
Db	421	CAGCCCTGTGATGCCAGCAGGGGGCCCCCGAACCCTGCACAGAGTCTCTCTCAGGCAGT	480
Qy	729	GluSerLysProLysAspGluLeu	736
Db	481	GAATCGAAGCCCAAGGATGAGCTA	504

Search completed: July 18, 2004, 23:26:20  
Job time : 6563 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 07:57:32 ; Search time 105 Seconds  
(without alignments)  
1980.523 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 3870  
Sequence: 1 MAVRALKLTLLAVVAAS.....PPEPAQESLSGSEKPKDEL 736

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseqp 29Jan04:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3870	100.0	736	3 AAB36392	Aab36392 Human tum
2	3861	99.8	736	5 ABP69129	Abp69129 Human pol
3	3846	99.4	736	4 AAB93215	Aab93215 Human pro
4	3846	99.4	736	4 AAB88373	Aab88373 Human mem
5	3846	99.4	736	5 AAU09861	Aau09861 Novel hum
6	3844	99.3	736	4 AAB93142	Aab93142 Human pro
7	3801	98.2	806	5 ABG66709	Abg66709 Human nov
8	3801	98.2	806	5 ABG66686	Abg66686 Human nov
9	3645	94.2	706	5 ABP69128	Abp69128 Human pol
10	3582	92.6	804	4 AAB94482	Aab94482 Human pro
11	3572	92.3	708	5 AAU09860	Aau09860 Novel hum
12	3316	85.7	747	3 AAB36393	Aab36393 Mouse tum
13	3316	85.7	747	7 ABR63160	Abr63160 Murine po
14	3115	80.5	592	4 AAE04244	Aae04244 Human gen
15	2384	61.6	542	3 AAB36394	Aab36394 Mouse tum
16	2147	55.5	408	5 ABP69127	Abp69127 Human pol
17	2065	53.4	400	5 ABG73582	Abg73582 Human leu
18	1911	49.4	363	3 AAB36391	Aab36391 Human tum
19	1800	46.5	359	4 AAE04200	Aae04200 Human gen
20	1800	46.5	359	5 ABG64502	Abg64502 Human alb
21	1751	45.2	708	6 ADA54992	Ada54992 Human pro
22	1747	45.1	708	7 ABR63159	Abr63159 Human pro
23	1722	44.5	337	4 AAE03156	Aae03156 Human gen
24	1473	38.1	736	6 AAE02322	Aae02322 Human GPC
25	1409.5	36.4	527	4 AAB93094	Aab93094 Human pro

26	1179.5	30.5	551	6 AAO30313	Aao30313 Human pro
27	1123	25.0	217	3 AAB42734	Ab42734 Human ORF
28	977	25.2	182	6 ABU70516	Abu70516 Human adi
29	906.5	23.4	282	7 ADB82712	Ad82712 Human pro
30	836.5	21.6	267	4 AAU17171	Aau17171 Novel sig
31	836.5	21.6	267	4 AAU87588	Aau87588 Novel cen
32	836.5	21.6	267	7 ADB93879	Ad93879 Human nov
33	790	20.4	227	4 ASG16396	Asg16396 Novel hum
34	657	17.0	173	4 AAE04217	Aae04217 Human gen
35	657	17.0	173	5 ABG64501	Abg64501 Human alb
36	612	15.8	219	4 AAU87297	Aau87297 Novel cen
37	572	14.9	401	4 AAB93152	Aab93152 Human pro
38	572	14.8	401	7 ADE60315	Ad60315 Human pro
39	572	14.8	401	7 ADE60327	Ad60327 Human pro
40	572	14.8	401	7 ADE60319	Ad60319 Human pro
41	572	14.8	401	7 ADE60323	Ad60323 Human pro
42	572	14.8	401	7 ADE59277	Ad59277 Human pro
43	567	14.7	401	2 AAU17219	Aau17219 Human sec
44	567	14.7	401	4 AAU38989	Aau38989 Human sec
45	567	14.7	401	5 ABB55698	Abb55698 Human pol

ALIGNMENTS

RESULT 1

AAB36392  
ID AAB36392 standard; protein; 736 AA.  
XX  
AC AAB36392;  
XX  
DT 27-FEB-2001 (first entry)  
XX  
DE Human tumour suppressor Gros1-S protein SEQ ID NO:4.  
XX  
KW Tumour suppressor; Gros1-L; Gros1-S; cell proliferation; regulation;  
XX cancer; cytostatic; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN WC2000065047-A1.  
XX  
PD 02-NOV-2000.  
XX  
PF 26-APR-2000; 2000WO-JF002731.  
XX  
PR 26-APR-1999; 99JP-00118806.  
XX  
(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
XX  
PI Wadhwa R, Sugihara T, Yoshida A;  
XX  
DR WPI; 2000-687340/67.  
XX  
N-PSDB; AAC64725.  
XX  
PT Full-length tumor suppressor genes encoding Gros1-L, Gros1-S and mouse  
PT homologs participating in regulation of cell proliferation, useful in  
PT development of preventives and remedies of cancer.  
XX  
PS Claim 1; Page 78-83; 114pp; Japanese.  
XX  
CC The present sequence represents the human tumour suppressor designated  
CC Gros1-S. Gros1-L and Gros1-S have cytosolic activity and can be used in  
CC gene therapy. Gros1-L and Gros1-S genes are useful in the development of  
CC drugs used to treat and prevent cancer  
XX  
SQ Sequence 736 AA;

Query Match 100.0%; Score 3870; DB 3; Length 736;

Best Local Similarity 100.0%; Pred.No. 0;

Matches 736; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAVRALKLTLLAVVAASQAQVESEAGWGMVTPDLLFAEGTAAYARGDPGVVLSMER 60

DB 1 MAVRAKLLTLLAVVAASQAQVESEAGWGWTPDLLFAEGTAAYARGDWPVGLSMER 60  
 QY 61 ALRSRAALRALRLRCRTCAADFFWELDPDMSPPAQASGAGALRDLSPFGGLLRRACL 120  
 DB 61 ALRSRAALRALRLRCRTCAADFFWELDPDMSPPAQASGAGALRDLSPFGGLLRRACL 120  
 QY 121 RRCLGPPAAHSLSEMELEFRKRSPYNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180  
 DB 121 RRCLGPPAAHSLSEMELEFRKRSPYNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180  
 QY 181 QNLDDYQTMGVKQADFKDLTQPHMQEFLGRLVLYSEEQPQAVPHLEALQYFVAYE 240  
 DB 181 QNLDDYQTMGVKQADFKDLTQPHMQEFLGRLVLYSEEQPQAVPHLEALQYFVAYE 240  
 QY 241 ECRALCEGPDYDGVNLYEYNADLFOAITDHYIOVLNCKQNCVTELASHPSREKPFDFL 300  
 DB 241 ECRALCEGPDYDGVNLYEYNADLFOAITDHYIOVLNCKQNCVTELASHPSREKPFDFL 300  
 QY 301 PSHNYLQFAYYNIQNTQAGECAKTYLLFFPNDVNMNQLAYAAAMLGEEHTRSIGPRE 360  
 DB 301 PSHNYLQFAYYNIQNTQAGECAKTYLLFFPNDVNMNQLAYAAAMLGEEHTRSIGPRE 360  
 QY 361 SAKEYRQSRSLLEKELLFFAYDVGIPFVDPDSWTPPEVPIKRLQEKOKSERETAVRISQE 420  
 DB 361 SAKEYRQSRSLLEKELLFFAYDVGIPFVDPDSWTPPEVPIKRLQEKOKSERETAVRISQE 420  
 QY 421 IGNLMKETETLVEBKTESLVSRLTREGGPLLVEGSLTWSKLLNGYQVWMDGVISD 480  
 DB 421 IGNLMKETETLVEBKTESLVSRLTREGGPLLVEGSLTWSKLLNGYQVWMDGVISD 480  
 QY 481 HECBELQRLTNVAATSGDVGRTQSPHTPNKFKYGVTVFVKALKLGQSGKVPLOSAHLYYN 540  
 DB 481 HECBELQRLTNVAATSGDVGRTQSPHTPNKFKYGVTVFVKALKLGQSGKVPLOSAHLYYN 540  
 QY 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPHVDNCILNAETLVC 600  
 DB 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPHVDNCILNAETLVC 600  
 QY 601 VKEPPAYTFRYSAILYLNGDFDGNFYFTLDAKTVAEVQPCGQRAVGFSGTENPHG 660  
 DB 601 VKEPPAYTFRYSAILYLNGDFDGNFYFTLDAKTVAEVQPCGQRAVGFSGTENPHG 660  
 QY 661 VKATRGORCAIALWFTLDPRHSERDRVQADDLVKMLFSPPEMDLSOEQLDAQGPPPEP 720  
 DB 661 VKATRGORCAIALWFTLDPRHSERDRVQADDLVKMLFSPPEMDLSOEQLDAQGPPPEP 720  
 QY 721 AQESLSGSESXPKDEL 736  
 DB 721 AQESLSGSESXPKDEL 736

RESULT 2  
 ABP69129 ID ABP69129 standard; protein; 736 AA.  
 AC AC  
 XX ABP69129;  
 DT 20-JAN-2003 (first entry)  
 XX Human polypeptide SEQ ID NO 1176.  
 DE Human; genome mapping; gene therapy; food supplement; virus; fungus;  
 DE cell-proliferative disorder; neurodegenerative disease; bacterial;  
 KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
 KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
 KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;  
 KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
 KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
 KW antiarthritic.  
 OS Homo sapiens.  
 XX

PN WO200270539-A2.  
 PD 12-SEP-2002.  
 XX 05-MAR-2002; 2002WO-US005095.  
 PF 05-MAR-2001; 2001US-00799451.  
 XX (HYSE-) HYSEQ INC.  
 PA Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
 PI Wehrman T, Wang J, Wang D, Drmanac RT;  
 XX WPI: 2002-759812/82.  
 DR N-PSDB; AB211346.  
 XX New polynucleotides comprising sequences assembled from expressed  
 PT sequence tags (ESTs), useful for treating cell-proliferative,  
 PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet  
 PT or coagulation disorders.  
 XX Claim 9; SEQ ID NO 1176; 1012pp + Sequence Listing; English.  
 PS The invention relates to an isolated polynucleotide (I) comprising a  
 CC nucleotide sequence selected from any of 948 sequences (AB211119-  
 CC AB212066) or their mature protein coding portion, active domain coding  
 CC protein or complementary sequences. The polynucleotides are useful for  
 CC identifying expressed genes or for physical mapping of human genome. The  
 CC encoded polypeptides (ASP68902-ASP69849) are useful as molecular weight  
 CC markers, as a food supplement, for generating antibodies, in medical  
 CC imaging, screening and diagnostic assays and for treating cell-  
 CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's  
 CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,  
 CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,  
 CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver  
 CC or lung fibrosis infections (bacterial, viral, fungal, parasitic),  
 CC arthritis, etc. Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 736 AA;

Query Match 99.8%; Score 3861; DB 5; Length 736;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 734; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MAVRAKLLTLLAVVAASQAQVESEAGWGWTPDLLFAEGTAAYARGDWPVGLSMER 60  
 DB 1 MAVRAKLLTLLAVVAASQAQVESEAGWGWTPDLLFAEGTAAYARGDWPVGLSMER 60  
 QY 61 ALRSRAALRALRLRCRTCAADFFWELDPDMSPPAQASGAGALRDLSPFGGLLRRACL 120  
 DB 61 ALRSRAALRALRLRCRTCAADFFWELDPDMSPPAQASGAGALRDLSPFGGLLRRACL 120  
 QY 121 RRCLGPPAAHSLSEMELEFRKRSPYNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180  
 DB 121 RRCLGPPAAHSLSEMELEFRKRSPYNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180  
 QY 181 QNLDDYQTMGVKQADFKDLTQPHMQEFLGRLVLYSEEQPQAVPHLEALQYFVAYE 240  
 DB 181 QNLDDYQTMGVKQADFKDLTQPHMQEFLGRLVLYSEEQPQAVPHLEALQYFVAYE 240  
 QY 241 ECRALCEGPDYDGVNLYEYNADLFOAITDHYIOVLNCKQNCVTELASHPSREKPFDFL 300  
 DB 241 ECRALCEGPDYDGVNLYEYNADLFOAITDHYIOVLNCKQNCVTELASHPSREKPFDFL 300  
 QY 301 PSHNYLQFAYYNIQNTQAGECAKTYLLFFPNDVNMNQLAYAAAMLGEEHTRSIGPRE 360  
 DB 301 PSHNYLQFAYYNIQNTQAGECAKTYLLFFPNDVNMNQLAYAAAMLGEEHTRSIGPRE 360  
 QY 361 SAKEYRQSRSLLEKELLFFAYDVGIPFVDPDSWTPPEVPIKRLQEKOKSERETAVRISQE 420

Db 361 SAKEYRQSRLLKELLFFAYDVDFGIFVDPDSHTPEVTPKRLQEKQKSERETAVRISOE 420  
QY 421 IGNLMKEIETLVBEKTKESLDVSGRLTREGGPLYEGISLTMSKLLNGYQRVVMDGVISD 480  
Db 421 IGNLMKEIETLVBEKTKESLDVSGRLTREGGPLYEGISLTMSKLLNGYQRVVMDGVISD 480  
QY 481 HECQELQRLTNVAATSGDGYRGQTSPTNEKEFGYGVTFKALKLGQEGKVPLOSAHLYN 540  
Db 481 HECQELQRLTNVAATSGDGYRGQTSPTNEKEFGYGVTFKALKLGQEGKVPLOSAHLYN 540  
QY 541 VTEKVRIMESYFRLDTPLYFSYSHLVCRTAIEEVOAERKDDSHPVHVNDNCILNAETLVC 600  
Db 541 VTEKVRIMESYFRLDTPLYFSYSHLVCRTAIEEVOAERKDDSHPVHVNDNCILNAETLVC 600  
QY 601 VKPPPAYTFRDYSAILYLNGDFDGGNFYFTLDAKTVAEVOQCRVGFSSGTENPHG 660  
Db 601 VKPPPAYTFRDYSAILYLNGDFDGGNFYFTLDAKTVAEVOQCRVGFSSGTENPHG 660  
QY 661 VKAVTGRGRCALIAFWTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEPDLDAQQQPPRP 720  
Db 661 VKAVTGRGRCALIAFWTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEPDLDAQQQPPRP 720  
QY 721 AQESLSGSESKPKDEL 736  
Db 721 AQESLSGSESKPKDEL 736

RESULT 3  
AAB93215  
ID AAB93215 standard; protein; 736 AA.  
AC AAB93215;  
XX  
DT 26-JUN-2001 (first entry)  
DE Human protein sequence SEQ ID NO:12194.  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX Homo sapiens.  
XX EPI074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-00116126.  
XX  
PR 29-JUL-1999; 99JP-00248036.  
PR 27-AUG-1999; 99JP-00300253.  
PR 11-JAN-2000; 2000JP-00118776.  
PR 02-MAY-2000; 2000JP-00183767.  
PR 09-JUN-2000; 2000JP-00241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
XX WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
XX  
PS Claim 8; SEQ ID NO 12194; 2537pp + Sequence Listing; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination

of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH15742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX  
SQ Sequence 736 AA;

Query Match 99.4%; Score 3846; DB 4; Length 736;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAVRALKLLTLLAVVAAASQAEESEAGWGMVTPDLLFAEGTAAYARGDWPVGLSMER 60  
Db 1 MAVRALKLLTLLAVVAAASQAEESEAGWGMVTPDLLFAEGTAAYARGDWPVGLSMER 60

QY 61 ALRSPALRALRLRCRTCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
Db 61 ALRSPALRALRLRCRTCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120

QY 121 RRLGPPAAHLSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180  
Db 121 RRLGPPAAHLSLSEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180

QY 181 QNLVYQTMGSKGKADFKDLETPHMQBFRGLVRLYSEEQPQEAHPHLEALQRYFYAYE 240  
Db 181 QNLVYQTMGSKGKADFKDLETPHMQBFRGLVRLYSEEQPQEAHPHLEALQRYFYAYE 240

QY 241 ECRALCEGPYDGYNYLYENADLFOAITDHYIQLVNCQKNCVTELASHPSREKPFEDFL 300  
Db 241 ECRALCEGPYDGYNYLYENADLFOAITDHYIQLVNCQKNCVTELASHPSREKPFEDFL 300

QY 301 PSHYNYLOFAYYNIQNTQAGECAKTYLLFFPNDEVMNQNLAYYAAMLGEEHTRSIGPRE 360  
Db 301 PSHYNYLOFAYYNIQNTQAGECAKTYLLFFPNDEVMNQNLAYYAAMLGEEHTRSIGPRE 360

QY 361 SAKEYRQSRLLKELLFFAYDVDFGIFVDPDSHTPEVTPKRLQEKQKSERETAVRISOE 420  
Db 361 SAKEYRQSRLLKELLFFAYDVDFGIFVDPDSHTPEVTPKRLQEKQKSERETAVRISOE 420

QY 421 IGNLMKEIETLVBEKTKESLDVSGRLTREGGPLYEGISLTMSKLLNGYQRVVMDGVISD 480  
Db 421 IGNLMKEIETLVBEKTKESLDVSGRLTREGGPLYEGISLTMSKLLNGYQRVVMDGVISD 480

QY 481 HECQELQRLTNVAATSGDGYRGQTSPTNEKEFGYGVTFKALKLGQEGKVPLOSAHLYN 540  
Db 481 HECQELQRLTNVAATSGDGYRGQTSPTNEKEFGYGVTFKALKLGQEGKVPLOSAHLYN 540

QY 541 VTEKVRIMESYFRLDTPLYFSYSHLVCRTAIEEVOAERKDDSHPVHVNDNCILNAETLVC 600  
Db 541 VTEKVRIMESYFRLDTPLYFSYSHLVCRTAIEEVOAERKDDSHPVHVNDNCILNAETLVC 600

QY 601 VKPPPAYTFRDYSAILYLNGDFDGGNFYFTLDAKTVAEVOQCRVGFSSGTENPHG 660  
Db 601 VKPPPAYTFRDYSAILYLNGDFDGGNFYFTLDAKTVAEVOQCRVGFSSGTENPHG 660

QY 661 VKAVTGRGRCALIAFWTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEPDLDAQQQPPRP 720  
Db 661 VKAVTGRGRCALIAFWTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEPDLDAQQQPPRP 720

QY 721 AQESLSGSESKPKDEL 736  
Db 721 AQESLSGSESKPKDEL 736

Db 721 AQESLSGSESKPKDEL 736

RESULT 4

AA883373

ID AAB88373 standard; protein; 736 AA.

XX AC AAB88373;

XX DT 23-MAY-2001 (first entry)

XX DE Human membrane or secretory protein clone PSEC0109.

XX KW Human; secretory protein; membrane protein; vaccine; gene therapy;

XX KW rheumatoid arthritis; diabetes.

XX OS Homo sapiens.

XX EN EPI067182-A2.

XX PD 10-JAN-2001.

XX PF 07-JUL-2000; 2000EP-00114090.

XX PR 08-JUL-1999; 99JP-00194179.

XX PR 11-JAN-2000; 2000JP-00118775.

XX PR 02-MAY-2000; 2000JP-00183766.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;

XX DR WPI; 2001-093989/11.

XX DR N-PSDB; AAF93800.

XX PT Nucleic acids encoding secretory proteins/membrane proteins, useful in

XX PT gene therapy or as candidate target molecules in drug development.

XX PS Claim 1; SEQ ID NO 114; 609pp + Sequence Listing; English.

XX CC This invention relates to nucleic acid sequences AAF93744 - AAF93916

XX CC which encode human secretory or membrane proteins represented by AAB88317

XX CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and

XX CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the

XX CC invention. The invention also includes methods for the production of

XX CC antibodies directed against the proteins, and cDNA sequences, which can

XX CC be used in vaccines. The polynucleotide sequences can be used in gene

XX CC therapy. The polynucleotide sequences and the proteins they encode may be

XX CC used in the prevention, treatment and diagnosis of diseases associated

XX CC with inappropriate secretory protein/membrane protein expression. The

XX CC nucleic acids and complementary sequences may also be used as DNA probes

XX CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect

XX CC and quantitate the presence of similar nucleic acid sequences in samples.

XX CC They may also be used to study the expression and function of secretory

XX CC proteins/membrane polypeptides and their role in metabolism. The

XX CC polypeptides may be used as antigens in the production of antibodies

XX CC against them and in assays to identify modulators (agonists and

XX CC antagonists) of expression and activity. The antibodies and antagonists

XX CC may also be used as therapeutic agents to down regulate expression and

XX CC activity. The antibodies may also be used as diagnostic agents for

XX CC detecting the presence of the polypeptides in samples (e.g. by enzyme

XX CC linked immunosorbent assay (ELISA). Examples of diseases which may be

XX CC treated include rheumatoid arthritis and diabetes

XX SQ Sequence 736 AA;

Query Match 99.4%; Score 3846; DB 4; Length 736;

Best Local Similarity 99.6%; Pred. No. 0;

Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MAVRALKLLTLLAVVAASQAQVESEAGWGVTPDLLFAEGTAAAYARGDWPGLWSMER 60

Db 1 MAVRALKLLTLLAVVAASQAQVESEAGWGVTPDLLFAEGTAAAYARGDWPGLWSMER 60

QY 61 ALRSRAALRALRLRCRTCAADFPWELDPDWSPPAQASGAGALRDLISFFGGLLRRACL 120

DB 61 ALRSRAALRALRLRCRTCAADFPWELDPDWSPPAQASGAGALRDLISFFGGLLRRACL 120

QY 121 RCLGPPAAHLSLEEMELEFRKRSPYNYLVAYFKINKLEKAVAAAHHTFFVGNBEHMEMQ 180

DB 121 RCLGPPAAHLSLEEMELEFRKRSPYNYLVAYFKINKLEKAVAAAHHTFFVGNBEHMEMQ 180

QY 181 QNLDDYQTMVGKADFKDLETQPHMQEFLGRLVLYSEEQPQEAHPHLEAALQYFVAYE 240

DB 181 QNLDDYQTMVGKADFKDLETQPHMQEFLGRLVLYSEEQPQEAHPHLEAALQYFVAYE 240

QY 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTTELASHPSREKPFEDFL 300

DB 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQVLNCKQNCVTTELASHPSREKPFEDFL 300

QY 301 PSHYNYLOFAYNYIGNYTQAGECAKTYLLFPNDEVMMONLAYAAMLGEETHSIGPRE 360

DB 301 PSHYNYLOFAYNYIGNYTQAGECAKTYLLFPNDEVMMONLAYAAMLGEETHSIGPRE 360

QY 361 SAKYRQSRSLLEKELLFFAYDVGIFPDPSWTPEEVI PKRLOEKQKSERETAVRISQE 420

DB 361 SAKYRQSRSLLEKELLFFAYDVGIFPDPSWTPEEVI PKRLOEKQKSERETAVRISQE 420

QY 421 IGNLMKEITLVEEKTESLDVSRLTREGGDLLEGISLTWNSKLLNGYQVWMDGVISD 480

DB 421 IGNLMKEITLVEEKTESLDVSRLTREGGDLLEGISLTWNSKLLNGYQVWMDGVISD 480

QY 481 HECQELQLRLTNVAATSGDYGQTSPTPNEKFGYGVTVFKALKLCQEGKVPLOSAHLYN 540

DB 481 HECQELQLRLTNVAATSGDYGQTSPTPNEKFGYGVTVFKALKLCQEGKVPLOSAHLYN 540

QY 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRVTAIEVQAEKDDSHPVHVDNCLNAETLVC 600

DB 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRVTAIEVQAEKDDSHPVHVDNCLNAETLVC 600

QY 601 VKEPPAYTFRDIYSALLYLNGDFDGNFYFTBELDKTVAEVQPOCGRAVGFSSGTENPHG 660

DB 601 VKEPPAYTFRDIYSALLYLNGDFDGNFYFTBELDKTVAEVQPOCGRAVGFSSGTENPHG 660

QY 661 VKAVTRGQRCALWFTLDRHSERDRVQADDLVKMLFSPPEMDLSQEPDLDAQGPPPEP 720

DB 661 VKAVTRGQRCALWFTLDRHSERDRVQADDLVKMLFSPPEMDLSQEPDLDAQGPPPEP 720

QY 721 AQESLSGSESKPKDEL 736

DB 721 AQESLSGSESKPKDEL 736

RESULT 5

AAU09861

ID AAU09861 standard; protein; 736 AA.

XX AC AAU09861;

XX DT 26-FEB-2002 (first entry)

XX DE Novel human secreted protein #2.

XX KW Secreted protein; cytostatic; immunosuppressive; vulnary; vaccine;

XX KW antinflammatory; neuroprotective; nephrotropic; cardiovascular; human;

XX KW cancer; autoimmune disease; wound healing disorder; infection;

XX KW haematopoietic disorder; inflammatory disorder; infertility;

XX KW neurological disease; psychiatric disease; cardiovascular disease;

XX KW respiratory disease; renal; gastrointestinal.

XX OS Homo sapiens.

XX FN WO200179454-A1.

XX PD 25-OCT-2001.

PF 11-APR-2001; 2001WO-US011797.  
XX 13-APR-2000; 2000US-0196603P.  
PR 24-APR-2000; 2000US-0199417P.  
XX (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
XX Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z;  
PI WPI; 2002-061975/08.  
DR N-PSDB; AAS17573.  
XX  
PT New secreted proteins or polypeptides, useful for treating e.g. cancer,  
PT autoimmune diseases, wound healing disorder, infections, hematopoietic  
PT disorders, inflammatory disorders, infertility, cancer.  
XX  
PS Claim 1; Page 54-56; 92pp; English.  
XX  
CC The invention relates to an isolated novel secreted polypeptide (I) and  
CC polynucleotide (II). (I) and (II) are useful for treating cancer,  
CC autoimmune diseases, wound healing disorder, infections, hematopoietic  
CC disorders, inflammatory disorders, infertility, neurological and  
CC psychiatric diseases, cardiovascular diseases, respiratory diseases,  
CC renal diseases, or gastrointestinal diseases. These may also be used to  
CC treat diseases, abnormalities and disorders caused by abnormal  
CC expression, production, function and/or metabolism of the genes, as  
CC vaccines for inducing immunological response in a mammal, and in  
CC screening methods for detecting the effect of added compounds on the  
CC production of mRNA and polypeptide in cells. The polypeptides can be used  
CC as immunogens to produce antibodies immunospecific for the polypeptides,  
CC and to identify membrane-bound or soluble receptors. The polynucleotides  
CC may be used as diagnostic reagents, in chromosome localisation studies,  
CC and in tissue expression studies. The present sequence represents the  
CC amino acid sequence of novel human secreted protein #2  
XX  
SQ Sequence 736 AA;  
Query Match 99.4%; Score 3846; DB 5; Length 736;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MAVRALKLTLLAVVAASQAESEAGGWMVTPDLLFAEGTAAYARGDWPVGLSMER 60  
Db 1 MAVRALKLTLLAVVAASQAESEAGGWMVTPDLLFAEGTAAYARGDWPVGLSMER 60  
QY 61 ALRSRALRALRLRCRTOCAADFPWELDPWSPQAASGAGALRDLSPFGGLLRRAACL 120  
Db 61 ALRSRALRALRLRCRTOCAADFPWELDPWSPQAASGAGALRDLSPFGGLLRRAACL 120  
QY 121 RRLGLPPAAHSLSEEMLEPRKSPYNYLQVAYFKINKLEKAVAAAHHTFFVGNPEHMQ 180  
Db 121 RRLGLPPAAHSLSEEMLEPRKSPYNYLQVAYFKINKLEKAVAAAHHTFFVGNPEHMQ 180  
QY 181 QNLDYQYTMNGVKEADFKDLETPHMOERFLGVRLYSEEQPQAVPHLEAALQYFVAYE 240  
Db 181 QNLDYQYTMNGVKEADFKDLETPHMOERFLGVRLYSEEQPQAVPHLEAALQYFVAYE 240  
QY 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQLVNLCKQNCVTELASHPSREKPPEDFL 300  
Db 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIQLVNLCKQNCVTELASHPSREKPPEDFL 300  
QY 301 PSHNYLQFAYNYIGNYTOAGECAKYLFLFPNDENVQNLAYYAAMLGEHSTRSGPRE 360  
Db 301 PSHNYLQFAYNYIGNYTOAGECAKYLFLFPNDENVQNLAYYAAMLGEHSTRSGPRE 360  
QY 361 SAKEXRQSLLEKELFFAYDVDFGIPDPDSDWTPPEVTPKRLQEKOKSERETAVRISOE 420  
Db 361 SAKEXRQSLLEKELFFAYDVDFGIPDPDSDWTPPEVTPKRLQEKOKSERETAVRISOE 420  
QY 421 IGLNLMKEIETLVEKTKESLDVSRRLTREGGPLLVEGSLTMSKLNQYQVVMQGVISD 480  
Db 421 IGLNLMKEIETLVEKTKESLDVSRRLTREGGPLLVEGSLTMSKLNQYQVVMQGVISD 480

QY 481 HECOEQLRLTNVAATSGDGYRGQTSPTTPNEKFGVTVFVKALKLGQSGKVPLOSAHLYYN 540  
Db 481 HECOEQLRLTNVAATSGDGYRGQTSPTTPNEKFGVTVFVKALKLGQSGKVPLOSAHLYYN 540  
QY 541 VTEKVRIMESYFRLDTPLYFSYSHLVCRTAIEVQAEKDDSHPVHVDNCILNAETLVC 600  
Db 541 VTEKVRIMESYFRLDTPLYFSYSHLVCRTAIEVQAEKDDSHPVHVDNCILNAETLVC 600  
QY 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTTDLAKTVTAEVQPCGGRVAGSSGTENPHG 660  
Db 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTTDLAKTVTAEVQPCGGRVAGSSGTENPHG 660  
QY 661 VKAVTRGORCAIALWFTLDPHRSERDRVQADDLVKMLFSPREMDLSOBQPLDAQGPPPEP 720  
Db 661 VKAVTRGORCAIALWFTLDPHRSERDRVQADDLVKMLFSPREMDLSOBQPLDAQGPPPEP 720  
QY 721 AQESLSGSESKPKDEL 736  
Db 721 AQESLSGSESKPKDEL 736  
RESULT 6  
AAB93142  
ID AAB93142 standard; protein; 736 AA.  
XX AAB93142;  
AC AAB93142;  
DT 26-JUN-2001 (first entry)  
XX Human protein sequence SEQ ID NO:12045.  
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
KW Homo sapiens.  
XX  
OS Homo sapiens.  
XX  
FN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-00116126.  
XX  
PR 29-JUL-1999; 99JP-00248036.  
PR 27-AUG-1999; 99JP-00300253.  
PR 11-JAN-2000; 2000JP-00118776.  
PR 02-MAY-2000; 2000JP-00183767.  
PR 09-JUN-2000; 2000JP-00241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
PT  
XX  
PS Claim 8; SEQ ID NO 12045; 2537pp + Sequence Listing; English.  
XX  
CC The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesising polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 XX  
 SQ Sequence 736 AA;

Query Match 99.3%; Score 3844; DB 4; Length 736;  
 Best Local Similarity 99.5%; Pred. No. 0;  
 Matches 732; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy	1	MAVRALKLLTTLAVAAASQAQVESEAGMGWVTPDLLFAEGTAAYARGDWPQVLSMER	60
Db	1	MAVRALKLLTTLAVAAASQAQVESEAGMGWVTPDLLFAEGTAAYARGDWPQVLSMER	60
Qy	61	ALRSRAALRALRLCRTOCAADPWELDPWSPPAQASGAGALRDLSPFGGLLRAACL	120
Db	61	ALRSRAALRALRLCRTOCAADPWELDPWSPPAQASGAGALRDLSPFGGLLRAACL	120
Qy	121	RRCLGPPAAHSLSEMELEPRKSPYNYLQVAFKINKLEKAVAAAHTFFVGNPEHMEQ	180
Db	121	RRCLGPPAAHSLSEMELEPRKSPYNYLQVAFKINKLEKAVAAAHTFFVGNPEHMEQ	180
Qy	181	QNLDDYQTMGSVKEADFKDLETQPHMQEFLGRLVLYSEEOQEAHPLEALQEYVAYE	240
Db	181	QNLDDYQTMGSVKEADFKDLETQPHMQEFLGRLVLYSEEOQEAHPLEALQEYVAYE	240
Qy	241	ECRALCEGPDYDGYNYLQVAFKINKLEKAVAAAHTFFVGNPEHMEQ	300
Db	241	ECRALCEGPDYDGYNYLQVAFKINKLEKAVAAAHTFFVGNPEHMEQ	300
Qy	301	PSHNYLQFAYNIGNYQTOGCAKTYLLFPFNDVNMNQNLAYYAMLGEEHTRSGPRE	360
Db	301	PSHNYLQFAYNIGNYQTOGCAKTYLLFPFNDVNMNQNLAYYAMLGEEHTRSGPRE	360
Qy	361	SAKEYRQSRSLLEKELLFPAYDVGIPFVDPDSWTPPEVTPKRLQEKQKSERETAVRISOE	420
Db	361	SAKEYRQSRSLLEKELLFPAYDVGIPFVDPDSWTPPEVTPKRLQEKQKSERETAVRISOE	420
Qy	421	IGNLMKEIETLVBKTESLDVSRLETREGGGLLYEGISLTMSKLNGLYQVVMQGVISD	480
Db	421	IGNLMKEIETLVBKTESLDVSRLETREGGGLLYEGISLTMSKLNGLYQVVMQGVISD	480
Qy	481	HECOELQRLTNVAATSGDVGRTSHTPNKEKFGVTVFKALKGQKGVKPLQSAHLYN	540
Db	481	HECOELQRLTNVAATSGDVGRTSHTPNKEKFGVTVFKALKGQKGVKPLQSAHLYN	540
Qy	541	VTEKVRIMESYFRDLTPFLYFSYSHLVCRITAEVQAEKRDSDSHPVHVDNCLNAETLVC	600
Db	541	VTEKVRIMESYFRDLTPFLYFSYSHLVCRITAEVQAEKRDSDSHPVHVDNCLNAETLVC	600
Qy	601	VKEPPATFRDYSAILYNGDFGGNPFYFTELLDAKTVTAEVQCCGRVCFSSGTENPHG	660
Db	601	VKEPPATFRDYSAILYNGDFGGNPFYFTELLDAKTVTAEVQCCGRVCFSSGTENPHG	660
Qy	661	VKAVTRQRCALALFWFLDPRHSERDRVQADDLVKMLFGPEEMDLQEOPLDAQQGFPP	720
Db	661	VKAVTRQRCALALFWFLDPRHSERDRVQADDLVKMLFGPEEMDLQEOPLDAQQGFPP	720
Qy	721	AQESLSGSESKPKDEL 736	
Db	721	AQESLSGSESKPKDEL 736	

ID	ABG66709 standard; protein; 806 AA.
XX	
AC	ABG66709;
XX	
DT	30-AUG-2002 (first entry)
XX	
DE	Human novel polypeptide #44.
XX	
KW	Human; inflammatory condition; shock; sepsis; immune response; cancer;
KW	wound healing; central nervous system disease; haematopoiesis;
KW	peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
KW	myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
KW	cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
KW	bone degenerative disorder; periodontal disease; reperfusion injury;
KW	lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
KW	allergic condition; thrombolysis; thrombosis; coagulation disorder;
KW	fungal infection.
XX	
OS	Homo sapiens.
XX	
PN	WO200244340-A2.
XX	
PD	06-JUN-2002.
XX	
PF	30-NOV-2001; 2001WO-US047004.
XX	
PR	30-NOV-2000; 2000US-00728952.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;
PI	Yamazaki V, Ujwal ML, Drmanac RT;
XX	
DR	WPI; 2002-508509/54.
XX	N-PSDB; ABK94933.
PT	Novel nucleic acids and polypeptides for diagnosis, treatment of
PT	inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
XX	disorders, cancer and promoting wound healing.
PS	Claim 10; Page 616-617; 672pp; English.
XX	
CC	The invention relates to human novel polynucleotides and associated
CC	polypeptides. The polynucleotides and polypeptides are useful for
CC	treating inflammatory conditions such as arthritis, nephritis, Crohn's
CC	disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
CC	and cancer and for promoting wound healing. The sequences are used to
CC	induce the proliferation of neural cells and regeneration of nerve and
CC	brain tissue, and are useful for the treatment of central and peripheral
CC	nervous system diseases and neuropathies, such as Alzheimer's disease,
CC	Parkinson's disease, Huntington's disease and amyotrophic lateral
CC	sclerosis. The sequences are involved in chemotactic or chemokinetic
CC	activity, regulation of haematopoiesis, treatment of myeloid or lymphoid
CC	cell disorders and platelet disorders such as thrombocytopenia,
CC	regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC	growth, tissue repair, healing of burns, incisions, ulcers, treatment of
CC	osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
CC	disease. The sequences of the invention are also useful for gut
CC	protection or regeneration and treatment of lung or liver fibrosis,
CC	reperfusion injury in various tissues, immune deficiencies and disorders
CC	including severe combined immunodeficiency (SCID), bacterial or fungal
CC	infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
CC	gravis, allergic conditions such as asthma, thrombolysis or thrombosis
CC	and coagulation disorders. Sequences ABG66666-ABG66758 represent human
CC	novel polypeptides of the invention
XX	
SQ	Sequence 806 AA;

Query Match 98.2%; Score 3801; DB 5; Length 806;  
 Best Local Similarity 90.9%; Pred. No. 0;  
 Matches 733; Conservative 0; Mismatches 3; Indels 70; Gaps 1;



Db 1 MAVRALKLLTLLAVAAASQAQVESEAGWMTPTDLLFAEGTAAYARGDWPVGLSMER 60  
Qy 61 ALRSRAALRALRLCRCTCCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
Db 61 ALRSRAALRALRLCRCTCCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
Qy 121 RRCGLPPAAHSLSEMELEFRKSPYNYLQVAYF-----KINKLEKAVAAAHFFF 154  
Db 121 RRCGLPPAAHSLSEMELEFRKSPYNYLQVAYF-----KINKLEKAVAAAHFFF 180  
Qy 155 -----KINKLEKAVAAAHFFF 170  
Db 181 GDRGVRREGKVASWLGSSPRSGELLPGRRPSPSSHGQMLTFKINKLEKAVAAAHFFF 240  
Qy 171 VGNPEHMEMQNLDYYQTMGSKVEADFKDLETPHMQEFLRGVRLYSEEQPQAVPHLEA 230  
Db 241 VGNPEHMEMQNLDYYQTMGSKVEADFKDLETPHMQEFLRGVRLYSEEQPQAVPHLEA 300  
Qy 231 ALQYFYVAYBECCALCEGPDYDGYNYLEYNADLFOAITDHYIQVLNCKNCVCUTELASHP 290  
Db 301 ALQYFYVAYBECCALCEGPDYDGYNYLEYNADLFOAITDHYIQVLNCKNCVCUTELASHP 360  
Qy 291 SREKPFEDFLPSHNYLQFAYNIGNYTOAGECAKTVLLFFPNDEVMNQNLAYYAMLGE 350  
Db 361 SREKPFEDFLPSHNYLQFAYNIGNYTOAGECAKTVLLFFPNDEVMNQNLAYYAMLGE 420  
Qy 351 EHTSICGPRSAKEYRQSRLLKELLFPAYDVFQIPVDPSWTPPEVPIKRLQEKQKSE 410  
Db 421 EHTSICGPRSAKEYRQSRLLKELLFPAYDVFQIPVDPSWTPPEVPIKRLQEKQKSE 480  
Qy 411 RETAVRISQEIGNLMEKIEITLVEKTKESLDVSRLTREGGPLLVEGISTLMTNSKLNGYQ 470  
Db 481 RETAVRISQEIGNLMEKIEITLVEKTKESLDVSRLTREGGPLLVEGISTLMTNSKLNGSQ 540  
Qy 471 RVMDGVISDEHCEQLRLTNVAATSGDGYRGQTSPTPNEKFGYVTVFKALXLGQEGKV 530  
Db 541 RVMDGVISDEHCEQLRLTNVAATSGDGYRGQTSPTPNEKFGYVTVFKALXLGQEGKV 600  
Qy 531 PLOSAHLYNVTXKVRIMESYFRLDTPLYFSYSHLVCRTAIEVQAEKRDSDHPVHVDN 590  
Db 601 PLOSAHLYNVTXKVRIMESYFRLDTPLYFSYSHLVCRTAIEVQAEKRDSDHPVHVDN 660  
Qy 591 CILNAETLVCKEPPATFDYSAIYLNGDFDGNFYFTLDAKTVTAEVQPCQRAVG 650  
Db 661 CILNAETLVCKEPPATFDYSAIYLNGDFDGNFYFTLDAKTVTAEVQPCQRAVG 720  
Qy 651 PSSGTENPHGVKAVTRGQRCALALWFTLDRHSEDRVQADDLVKMLFSPSEMDLSQEQP 710  
Db 721 PSSGTENPHGVKAVTRGQRCALALWFTLDRHSEDRVQADDLVKMLFSPSEMDLSQEQP 780  
Qy 711 LDACGGPPEPAQESLSGSESKPKDEL 736  
Db 781 LDACGGPPEPAQESLSGSESKPKDEL 806  
RESULT 8  
ID ABG66686 standard; protein; 806 AA.  
XX AC ABG66686;  
XX DT 30-AUG-2002 (first entry)  
XX DE Human novel polypeptide #21.  
XX KW Human; inflammatory condition; shock; sepsis; immune response; cancer;  
KW wound healing; central nervous system disease; haematopoiesis;  
KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;  
KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;  
KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;  
KW bone degenerative disorder; periodontal disease; reperfusion injury;  
KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;

KW allergic condition; thrombolysis; thrombosis; coagulation disorder;  
KW fungal infection.  
XX OS Homo sapiens.  
PN WO200244340-A2.  
XX PD 06-JUN-2002.  
XX PF 30-NOV-2001; 2001WO-US047004.  
XX PR 30-NOV-2000; 2000US-00728952.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;  
PI Yamazaki V, Ujwal ML, Drmanac RT;  
XX DR WPI: 2002-508509/54.  
DR N-PSDB; ABX94910.  
XX PT Novel nucleic acids and polypeptides for diagnosis, treatment of  
PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell  
PT disorders, cancer and promoting wound healing.  
XX PS Claim 10; Page 584-585; 672pp; English.  
XX CC The invention relates to human novel polynucleotides and associated  
CC polypeptides. The polynucleotides and polypeptides are useful for  
CC treating inflammatory conditions such as arthritis, nephritis, Crohn's  
CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses  
CC and cancer and for promoting wound healing. The sequences are used to  
CC induce the proliferation of neural cells and regeneration of nerve and  
CC brain tissue, and are useful for the treatment of central and peripheral  
CC nervous system diseases and neuropathies, such as Alzheimer's disease,  
CC Parkinson's disease, Huntington's disease and amyotrophic lateral  
CC sclerosis. The sequences are involved in chemotactic or chemokinetic  
CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid  
CC cell disorders and platelet disorders such as thrombocytopenia,  
CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue  
CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of  
CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal  
CC disease. The sequences of the invention are also useful for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues, immune deficiencies and disorders  
CC including severe combined immunodeficiency (SCID), bacterial or fungal  
CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia  
CC gravis, allergic conditions such as asthma, thrombolysis or thrombosis  
CC and coagulation disorders. Sequences ABG66666-ABG66758 represent human  
CC novel polypeptides of the invention  
XX SQ Sequence 806 AA;  
Query Match 98.2%; Score 3801; DB 5; Length 806;  
Best Local Similarity 90.9%; Pred. No. 0;  
Matches 733; Conservative 0; Mismatches 3; Indels 70; Gaps 1;

Qy 1 MAVRALKLLTLLAVAAASQAQVESEAGWMTPTDLLFAEGTAAYARGDWPVGLSMER 60  
Db 1 MAVRALKLLTLLAVAAASQAQVESEAGWMTPTDLLFAEGTAAYARGDWPVGLSMER 60  
Qy 61 ALRSRAALRALRLCRCTCCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
Db 61 ALRSRAALRALRLCRCTCCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
Qy 121 RRCGLPPAAHSLSEMELEFRKSPYNYLQVAYF-----KINKLEKAVAAAHFFF 154  
Db 121 RRCGLPPAAHSLSEMELEFRKSPYNYLQVAYF-----KINKLEKAVAAAHFFF 180  
Qy 155 -----KINKLEKAVAAAHFFF 170  
Db 181 GDRGVRREGKVASWLGSSPRSGELLPGRRPSPSSHGQMLTFKINKLEKAVAAAHFFF 240

QY 171 VGNPEHMEQNLDDYYQTMGSKVEADFKDLTQPHMQEFLRGVLYSEBQFQAVPHLEA 230  
DB 241 VGNPEHMEQNLDDYYQTMGSKVEADFKDLTQPHMQEFLRGVLYSEBQFQAVPHLEA 300  
QY 231 ALQSYFVAYEBCRALCEGPDYDGYNYLYNADLFOAITDHYIOVLNCKQNCVTELASHP 290  
DB 301 ALQSYFVAYEBCRALCEGPDYDGYNYLYNADLFOAITDHYIOVLNCKQNCVTELASHP 360  
QY 291 SREKPFDFLPSHYNLYQFAYNIGNYNTQAGECAKTYLLFFPNDEVNQNLAAYAAMLGE 350  
DB 361 SREKPFDFLPSHYNLYQFAYNIGNYNTQAGECAKTYLLFFPNDEVNQNLAAYAAMLGE 420  
QY 351 EHTRSIGPRESAKYRORSLLLEKLLFFAYDVGIPFVDPDSWTPPEVPIKRLQEKQSE 410  
DB 421 EHTRSIGPRESAKYRORSLLLEKLLFFAYDVGIPFVDPDSWTPPEVPIKRLQEKQSE 480  
QY 411 RETAVRISQELGNLQKIEITLVEKTKESLDVSRLTREGGPLYEGISLTWNSKLLNGYQ 470  
DB 481 RETAVRISQELGNLQKIEITLVEKTKESLDVSRLTREGGPLYEGISLTWNSKLLNGSQ 540  
QY 471 RVMDGVISDHECQELQRLTNVAATSGDGYRGQTSPTPNEKFGYVTVFKALKLQGEQKV 530  
DB 541 RVMDGVISDHECQELQRLTNVAATSGDGYRGQTSPTPNEKFGYVTVFKALKLQGEQKV 600  
QY 531 PLQSAHLYNNTKVRIMESYFRDLTPLYPSYSHLVCRTAIEVQAEKDDSHPVHVDN 590  
DB 601 PLQSAHLYNNTKVRIMESYFRDLTPLYPSYSHLVCRTAIEVQAEKDDSHPVHVDN 660  
QY 591 CILNAETLVCVKEPPAYTFRDYSAILYNGDFDGNFYFTLDAKTVAEVPQCGRAVG 650  
DB 661 CILNAETLVCVKEPPAYTFRDYSAILYNGDFDGNFYFTLDAKTVAEVPQCGRAVG 720  
QY 651 PSSGTENPHGVKATRGORCAIALWFTLDRHSERDRVQADDLVKMLFSPSEMDLSQEQP 710  
DB 721 PSSGTENPHGVKATRGORCAIALWFTLDRHSERDRVQADDLVKMLFSPSEMDLSQEQP 780  
QY 711 LDAQGPPEPAQESLSGSESKPKDEL 736  
DB 781 LDAQGPPEPAQESLSGSESKPKDEL 806  
RESULT 9  
ID ABP69128  
XX ABP69128 standard; protein; 706 AA.  
AC ABP69128;  
XX  
DT 20-JAN-2003 (first entry)  
DE Human polypeptide SEQ ID NO 1175.  
KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
KW arthritis; cytostatic; immunomodulator; nontropic; dermatologic;  
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoasidic;  
KW antiarthritic.  
XX  
OS Homo sapiens.  
XX WO200270539-A2.  
XX  
PD 12-SEP-2002.  
XX  
PF 05-MAR-2002; 2002WO-US005095.  
XX  
PR 05-MAR-2001; 2001US-00799451.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Zhou P, Goodrich KW, Asundi V, Zhang J, Zhao QA, Ren P;

PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
PI Wehrman T, Wang J, Wang D, Drmanac RT;  
XX WPI: 2002-759812/82.  
DR N-PSDB; AB211345.  
XX  
PT New polynucleotides comprising sequences assembled from expressed  
PT sequence tags (ESTs), useful for treating cell-proliferative,  
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet  
PT or coagulation disorders.  
XX  
PS Claim 9; SEQ ID NO 1175; 1012pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated polynucleotide (1) comprising a  
CC nucleotide sequence selected from any of 948 sequences (AB211119-  
CC AB212066) or their mature protein coding portion, active domain coding  
CC protein or complementary sequences. The polynucleotides are useful for  
CC identifying expressed genes or for physical mapping of human genome. The  
CC encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight  
CC markers, as a food supplement, for generating antibodies, in medical  
CC imaging, screening and diagnostic assays and for treating cell-  
CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's  
CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,  
CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,  
CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver  
CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),  
CC arthritis, etc. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 706 AA;  
Query Match 94.2%; Score 3645; DB 5; Length 706;  
Best Local Similarity 95.0%; Pred. No. 1.1e-312;  
Matches 699; Conservative 2; Mismatches 5; Indels 30; Gaps 1;  
QY 1 MAVRALKLLTTLAVVAAASQAEVSEAGWGWTPDLLFAEGTAAAYARGDWGVLSMER 60  
DB 1 MAVRALKLLTTLAVVAAASQAEVSEAGWGWTPDLLFAEGTAAAYARGDWGVLSMER 60  
QY 61 ALRRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLISFFGLLRAACL 120  
DB 61 ALRRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLISFFGLLRAACL 120  
QY 121 RRLCGLPPAAHSLSEMELEFRKSPYNYLVQAYFKINKLEKAVAAAHFTFFVGNPEHMEQ 180  
DB 121 RRLCGLPPAAHSLSEMELEFRKSPYNYLVQAYFKINKLEKAVAAAHFTFFVGNPEHMEQ 180  
QY 181 QNLDDYYQTMGSKVEADFKDLTQPHMQEFLRGVLYSEBQFQAVPHLEAALQYFVAYE 240  
DB 181 QNLDDYYQTMGSKVEADFKDLTQPHMQEFLRGVLYSEBQFQAVPHLEAALQYFVAYE 240  
QY 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIOVLNCKQNCVTELASHPSREKPFDFDL 300  
DB 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIOVLNCKQNCVTELASHPSREKPFDFDL 300  
QY 301 PSHYNLYQFAYNIGNYNTQAGECAKTYLLFFPNDEVNQNLAAYAAMLGEHTSISIGRE 360  
DB 301 PSHYNLYQFAYNIGNYNTQAGECAKTYLLFFPNDEVNQNLAAYAAMLGEHTSISIGRE 360  
QY 361 SNAKEYRQSLLEKLLFFAYDVGIPFVDPDSWTPPEVPIKRLQEKQSERETAVRISQE 420  
DB 361 SNAKEYRQSLLEKLLFFAYDVGIPFVDPDSWTPPEVPIKRLQEKQSERETAVRISQE 480  
QY 421 IGLNLMKEIETLVEEKTESLDVSRLTREGGPLYEGISLTWNSKLLNGYQRVVMDGVISD 480  
DB 391 IGLNLMKEIETLVEEKTESLDVSRLTREGGPLYEGISLTWNSKLLNGYQRVVMDGVISD 450  
QY 481 HECQELQRLTNVAATSGDGYRGQTSPTPNEKFGYVTVFKALKLQGEQKVPLQSAHLYN 540  
DB 451 HECQELQRLTNVAATSGDGYRGQTSPTPNEKFGYVTVFKALKLQGEQKVPLQSAHLYN 510  
QY 541 VTEKVRIMESYFRDLTPLYPSYSHLVCRTAIEVQAEKDDSHPVHVDNLCILNAETLVC 600

Db 511 VTEKRRIMESYFRLDTPLYFSYSHLVCRITAEVQAERKDDSHPVHVDNCILNAETLVC 570  
Qy 601 VKPPPAYTFRDYSAIILYNGDFGGHFFFTLDAKTVTAHVQPCGRVGFSSGTENPHG 660  
Db 571 VKPPPAYTFRDYSAIILYNGDFGGHFFFTLDAKTVTAHVQPCGRVGFSSGTENPHG 630  
Qy 661 VKAVTRGQRCALWFTLDRHSDRVRQADDLVKMLFSPPEMDLSQEPDLAQOQPPPP 720  
Db 631 VKAVTRGQRCALWFTLDRHSDRVRQADDLVKMLFSPPEMDLSQEPDLAQOQPPPP 690  
Qy 721 AQESLSGSESKPKDEL 736  
Db 691 AQESLSGSESKPKDEL 706

RESULT 10  
AAB94482  
ID AAB94482 standard; protein; 804 AA.  
XX AAB94482;  
AC  
XX  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human protein sequence SEQ ID NO:15160.  
XX  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX  
XX Homo sapiens.  
XX  
PN EF1074617-A2.  
XX  
XX  
PD 07-FEB-2001.  
XX  
XX  
PF 28-JUL-2000; 2000EP-00116126.  
XX  
PR 29-JUL-1999; 99JP-00248036.  
PR 27-AUG-1999; 99JP-00300253.  
PR 11-JAN-2000; 2000JP-00118776.  
PR 02-MAY-2000; 2000JP-00183767.  
PR 03-JUN-2000; 2000JP-00241899.  
XX  
XX  
PA (HELI-) HELIX RES INST.

XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
XX WPI; 2001-3-8749/34.  
XX  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
XX  
XX  
PS Claim 8; SEQ ID NO 15160; 2537pp + Sequence Listing; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention  
XX  
SQ Sequence 804 AA;  
Query Match 92.6%; Score 3582; DB 4; Length 804;  
Best Local Similarity 99.3%; Pred.No.5.2e-307;  
Matches 682; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 MAVRALKLLTLLAVVAAASQAEESEAGWGVTPTDLLFABGTAYAYAGDWPVGVLSMER 60  
Db 1 MAVRALKLLTLLAVVAAASQAEESEAGWGVTPTDLLFABGTAYAYAGDWPVGVLSMER 60  
Qy 61 ALRSRAALRALRLRCRTQCAADFPWELDPDPSPAQASGAGALRDLISFFGGLLRAACL 120  
Db 61 ALRSRAALRALRLRCRTQCAADFPWELDPDPSPAQASGAGALRDLISFFGGLLRAACL 120  
Qy 121 RRCGLPPAAHLSSEMELEFRKRSPYNYLQVAYFKINKLEKAVAAAHFFVGNPEHMEMQ 180  
Db 121 RRCGLPPAAHLSSEMELEFRKRSPYNYLQVAYFKINKLEKAVAAAHFFVGNPEHMEMQ 180  
Qy 181 QNLDYYQTMGVKAEADFKDLETPHMQEFLRGVRLYSEEQPOEAVPHLEAALQYFYAYE 240  
Db 181 QNLDYYQTMGVKAEADFKDLETPHMQEFLRGVRLYSEEQPOEAVPHLEAALQYFYAYE 240  
Qy 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIOVLNCKQNCVTELASHPSREKPFEDFL 300  
Db 241 ECRALCEGPDYDGYNYLYNADLFOAITDHYIOVLNCKQNCVTELASHPSREKPFEDFL 300  
Qy 301 PSHYNYLQFAYYNYGTYQAGECAKTYLLFFPNDEVMNQNLAYYAAAMLGEHTRSIGPRE 360  
Db 301 PSHYNYLQFAYYNYGTYQAGECAKTYLLFFPNDEVMNQNLAYYAAAMLGEHTRSIGPRE 360  
Qy 361 SAKYRQSRSLLEKELLFPAYDVFGIPFVDPDSWTPPEVI PKLOEKQKSERETAVRISQE 420  
Db 361 SAKYRQSRSLLEKELLFPAYDVFGIPFVDPDSWTPPEVI PKLOEKQKSERETAVRISQE 420  
Qy 421 IGLNLMKEIETLVEEKTKESSLVSRITREGGPLLVEGISTWNKLLNGYQVWMDGVISD 480  
Db 421 IGLNLMKEIETLVEEKTKESSLVSRITREGGPLLVEGISTWNKLLNGYQVWMDGVISD 480  
Qy 481 HECQELQRLTNVAATSGDYGRTSPHTPNKFKYGVTVFKALKLGQEGKVPQLQSAHLYN 540  
Db 481 HECQELQRLTNVAATSGDYGRTSPHTPNKFKYGVTVFKALKLGQEGKVPQLQSAHLYN 540  
Qy 541 VTEKVRRTMESYFRLDTPLYFSYSHLVCRITAEVQAERKDDSHPVHVDNCILNAETLVC 600  
Db 541 VTEKVRRTMESYFRLDTPLYFSYSHLVCRITAEVQAERKDDSHPVHVDNCILNAETLVC 600  
Qy 601 VKEPPPAYTFRDYSAIILYNGDFDGGNFYFTLDAKTVTAHVQPCGRVGFSSGTENPHG 660  
Db 601 VKEPPPAYTFRDYSAIILYNGDFDGGNFYFTLDAKTVTAHVQPCGRVGFSSGTENPHG 660  
Qy 661 VKAVTRGQRCALWFTLDRHSDRVR 687  
Db 661 VKAVTRGQRCALWFTLDRHSDRVR 687

RESULT 11  
AAU09860  
ID AAU09860 standard; protein; 708 AA.  
XX  
AC AAU09860;  
XX  
DT 26-FEB-2002 (first entry)  
XX  
XX Novel human secreted protein #1.  
DE  
XX Secreted protein; cytostatic; immunosuppressive; vulnerary; vaccine; antinflammatory; neuroprotective; nephrotropic; cardiovascular; human;  
KW

KW cancer; autoimmune disease; wound healing disorder; infection;  
 KW haematopoietic disorder; inflammatory disorder; infertility;  
 KW neurological disease; psychiatric disease; cardiovascular disease;  
 KW respiratory disease; renal; gastrointestinal.

XX Homo sapiens.

PN WO200179454-A1.

XX 25-OCT-2001.

XX 11-APR-2001; 2001WO-US011797.

XX 13-APR-2000; 2000US-0196603P.

PR 24-APR-2000; 2000US-0199417P.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

PA (SMIK ) SMITHKLINE BEECHAM PLC.

XX Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z;

XX WPI; 2002-061975/08.

DR N-PSDB; AAS17572.

XX New secreted proteins or polypeptides, useful for treating e.g. cancer,  
 PT autoimmune diseases, wound healing disorder, infections, hematopoietic  
 PT disorders, inflammatory disorders, infertility, cancer.

XX Claim 1; Page 52-54; 92pp; English.

XX The invention relates to an isolated novel secreted polypeptide (I) and  
 CC polynucleotide (II). (I) and (II) are useful for treating cancer,  
 CC autoimmune diseases, wound healing disorder, infections, haematopoietic  
 CC disorders, inflammatory disorders, infertility, neurological and  
 CC psychiatric diseases, cardiovascular diseases, respiratory diseases,  
 CC renal diseases, or gastrointestinal diseases. These may also be used to  
 CC treat diseases, abnormalities and disorders caused by abnormal  
 CC expression, production, function and/or metabolism of the genes, as  
 CC vaccines for inducing immunological response in a mammal, and in  
 CC screening methods for detecting the effect of added compounds on the  
 CC production of mRNA and polypeptide in cells. The polypeptides can be used  
 CC as immunogens to produce antibodies immunospecific for the polypeptides,  
 CC and to identify membrane-bound or soluble receptors. The polynucleotides  
 CC may be used as diagnostic reagents, in chromosome localisation studies,  
 CC and in tissue expression studies. The present sequence represents the  
 CC amino acid sequence of novel human secreted protein #1

XX Sequence 708 AA;

Query Match 92.3%; Score 3572; DB 5; Length 708;  
 Best Local Similarity 93.9%; Pred. No. 3.2e-306;  
 Matches 691; Conservative 5; Mismatches 12; Indels 28; Gaps 3;  
 QY 1 MAVRALKLLTLLAVAAASQAQVESBAGWMTDILLFAEGTAAYARGDWPVGLSMER 60  
 DB 1 MAVRALKLLTLLAVAAASQAQVESBAGWMTDILLFAEGTAAYARGDWPVGLSMER 60  
 QY 61 ALRSGRAALRALRLCRTOCAADFPWELDPWSPSPAQASGAGALRDLSPFGGLLRAACL 120  
 DB 61 ALRSGRAALRALRLCRTOCAADFPWELDPWSPSPAQASGAGALRDLSPFGGLLRAACL 120  
 QY 121 RRCIGPPAAHSLSEMELEFRKSPYNYLOVAFKINKLEKAVAAAHFTFVGNPENEMQ 180  
 DB 121 RRCIGPPAAHSLSEMELEFRKSPYNYLOVAFKINKLEKAVAAAHFTFVGNPENEMQ 180  
 QY 181 QNLDDYQTMGSKVKEADFKDLTQPHMQEFLRGVRLYSEEQPQAVPHLEALQEFVAYE 240  
 DB 181 QNLDDYQTMGSKVKEADFKDLTQPHMQEFLRGVRLYSEEQPQAVPHLEALQEFVAYE 240  
 QY 241 ECRALCGPYDGYNYLYENADLFQAITDHYIQVLNCKNCVTELASHPSREKPFDEL 300  
 DB 241 ECRALCGPYDGYNYLYENADLFQAITDHYIQVLNCKNCVTELASHPSREKPFDEL 300

QY 301 PSHYNYLOFAYYNYGNITQAGECAKTYLLFPFNDEVNQNLAAYAAMLGEBHTRSIGPRE 360  
 DB 301 PSHYNYLOFAYYNYGNITQAGECAKTYLLFPFNDEVNQNLAAYAAMLGEBHTRSIGPRE 360  
 QY 361 SAKYRQSRSLLEKELLFFAYDVFGIPFVDPDSWTPPEEVIKRLQEKQKSERETAVRISQE 420  
 DB 361 SAKYRQSRSLLEKELLFFAYDVFGIPFVDPDSWTPPEEVIKRLQEKQKSERETAVRISQE 420  
 QY 421 IGNLMKEIETLVEEKTESLDVSRITREGGPLLREGISLTWNSKLNGYQVWMDGVISD 480  
 DB 421 IGNLMKEIETLVEEKTESLDVSRITREGGPLLREGISLTWNSKLNGYQVWMDGVISD 480  
 QY 481 HECQELQRLTNVAATSGDYGRTSPHTPNEKFGVTVFVKALKLGQEGKVPLOSAHLYYN 540  
 DB 481 HECQELQRLTNVAATSGDYGRTSPHTPNEKFGVTVFVKALKLGQEGKVPLOSAHLYYN 540  
 QY 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRITAEVQAEKDDSHPHVNDNCILNAETLVC 600  
 DB 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRITAEVQAEKDDSHPHVNDNCILNAETLVC 600  
 QY 601 VKEPPAYTFRDYSAITYLNGDFDGNFVFTELDAAKTVTAEVQPCGQRAVGFSSGTENPHG 660  
 DB 601 VKEPPAYTFRDYSAITYLNGDFDGNFVFTELDAAKTVTAEVQPCGQRAVGFSSGTENPHG 660  
 QY 661 VKAVTRGQRCALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQBPDLDAQGPPEP 720  
 DB 661 VKAVTRGQRCALWFTLDPHRSERDRVQADDLVKMLFSPPEMDLSQBPDLDAQGPPEP 720  
 QY 721 AQESLSGSSEKPKDEL 736  
 DB 721 AQESLSGSSEKPKDEL 736  
 QY 736 AQESLSGSSEKPKDEL 708  
 DB 736 AQESLSGSSEKPKDEL 708

# RESULT 12

AAB36393

ID AAB36393 standard; protein; 747 AA.

AC AAB36393;

DT 27-FEB-2001 (first entry)

DE Mouse tumour suppressor Gros1-L protein SEQ ID NO:6.

XX Tumour suppressor; Gros1-L; Gros1-S; cell proliferation; regulation;  
 KW cancer; cytostatic; gene therapy.

XX Mus musculus.

PN WO200065047-A1.

XX 02-NOV-2000.

XX 26-APR-2000; 2000WO-JP002731.

XX 26-APR-1999; 99JP-00118806.

PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX Wadhwa R, Sugihara T, Yoshida A;

XX WPI; 2000-687340/67.

XX N-PSDB; AAC64726.

PT Full-length tumor suppressor genes encoding Gros1-L, Gros1-S and mouse  
 PT homologs participating in regulation of cell proliferation, useful in  
 PT development of preventives and remedies of cancer.

XX Claim 1; Page 91-97; 114pp; Japanese.

CC The present sequence represents the mouse tumour suppressor designated  
 CC Gros1-L. Gros1-L and Gros1-S have cytostatic activity and can be used in  
 CC gene therapy. Gros1-L and Gros1-S genes are useful in the development of  
 CC drugs used to treat and prevent cancer

XX	Sequence	747 AA;
XX	Query Match	85.7%; Score 3316; DB 3; Length 747;
XX	Best Local Similarity	88.2%; Pred. No. 1.5e-283;
XX	Matches	623; Conservative 33; Mismatches 48; Indels 2; Gaps 1
QY	25	ESSEAGWGVYTDLLFAECTAAYARGDHPGVVLSVERALSRRAALRALRLRRCCTQCAADFP 84
DB	25	ESSEAGWGVYTDLLFAECTAAYARGDHPGVVLSVERALSRRAALRALRLRRCCTQCAADFP 84
QY	85	WELDDPDWSPSP--AAQASGAGALRDLSPFGGLLRAACLRCLGPPAAHSLSEEMELEFRK 142
DB	85	WAPDLDLDPDPSLSDPCGAAALHDLRFFGAVLRRAACLRCLGPPSAHLLSEELDLEFNK 144
QY	143	RSPPNYLOVAYPKINKLEKAVAAAHITFVGNPHEMEMQONLDYYQTMVGKAEKFDLET 202
DB	145	RSPPNYLOVAYFKINKLEKAVAAAHITFVGNPHEMEMQONLDYYQTMVGKAEKFDLEA 204
QY	203	QPHMOEFLRGLVLYSEEQPOEAVPHLSAALQEVFVAEECRALCEGPDYDGVNLYEYNA 262
DB	205	KPHMEHFLRGLVLYSEEQPOEAVPHLSAALQEVFVADECRALCEGPDYDGVNLYDVA 264
QY	263	DLFOAITDHYIQVLNCKQNCVTELASHSPREKPFDPFLPSHYNYLOFAYNIGNVTOAGE 322
DB	265	DLFOAITDHYIQVLNCKQNCVTELASHSPREKPFDPFLPSHYNYLOFAYNIGNVTOAIE 324
QY	323	CAKTYLLSFPNDENVNQNLAYAAMIGEEHTRSIGPRESAKEYQRSLAEKELLFPAYDV 382
DB	325	CAKTYLLSFPNDENVNQNLAYAAMIGEEHTRSIGPRESAKEYQRSLAEKELLFPAYDI 384
QY	383	FGIPVDPDPSWTPEEVIKRLQEKQKSERETAVRISOEIGNLMKEIETILVEEKTESLDV 442
DB	385	FGIPVDPDPSWTPEEVIKRLQEKQKSERETAVRISOEIGNLMKEIETILVEEKTESLDV 444
QY	443	SSLTREGGLLYEGISLITMNSKLVNGVQVVMOGVTSDEHCQELQSLTNVAATSGDGYRG 502
DB	445	SSLTREGGLLYEGISLITMNSKLVNGVQVVMOGVTSDEHCQELQSLTNVAATSGDGYRG 504
QY	503	QTSPTHTNEKFGVTVFKALKLGOEGKVPLOSAHLVYNVTEKVRVRWESYFLDTPLYFS 562
DB	505	QTSPTHTNEKFGVTVFKALKLGOEGKVPLOSAHLYNINTEKVRVRWESYFLDTPLYFS 564
QY	563	YSHLVCRTAIBEVOAERKDDSHPVHVDNCILNAFTLVCKPEPPAYTFRDYSAILYINGDF 622
DB	565	YSHFVCRTAIBESQAERKDSHPHVVDNCILNAEAFMCIKEPPAYTFREYSAILYINGDF 624
QY	623	DCGNFYFTELDKATTVTAESVQCCGRAVGFSGGTENPHGVKAVTRGQRCALMFTLDPRH 682
DB	625	DCGNFYFTELDKATTVTAESVQCCGRAVGFSGGTENPHGVKAVTRGQRCALMFTLDPRH 684
QY	683	SERDRVQADDLVKMLFSPPEMDLSQEQFLDAQGPPPEPAQESLGS 728
DB	685	SERDRVQADDLVKMLFSPPEEVDLPQEQFLPDQQGSPEEGEFLHGA 730
XX	RESULT 13	
XX	ABR63160	
XX	ABR63160 standard; protein; 747 AA.	
XX	ABR63160;	
XX	18-DEC-2003 (first entry)	
XX	Murine polypeptide.	
XX	Mouse; prollyhydroxylase-like protein; enzyme; cytostatic; cardiant;	
XX	antisthmatic; neuroleptic; antiparkinsonian; neuroprotective; nootropic;	
XX	uropathic; gene therapy.	
XX	Mus musculus.	
XX	WO2003066862-A1.	

XX	14-AUG-2003.
PD	
XX	
XX	04-FEB-2003; 2003WO-EP001082.
PF	
XX	
XX	04-FEB-2002; 2002US-0353190P.
PR	
PR	09-MAY-2002; 2002US-0378674P.
PA	(FARB ) BAYER AG.
XX	
XX	Smith TJ;
PI	
XX	
XX	WPI; 2003-663597/62.
DR	
XX	
XX	New polynucleotide and its encoded prollylhydroxylase-like protein useful
PT	for identifying modulators of protein activity and in gene therapy for
PT	treating disorders, e.g. cancer, Alzheimer's disease and genitourinary
PT	disorders.
XX	
XX	Disclosure; Page 121-123; 135pp; English.
PS	
XX	
CC	The present sequence is the protein sequence of a murine polypeptide. The
CC	invention relates to human prollylhydroxylase-like protein (see ABR63159),
CC	which is expressed in microvascular endothelial cells, prostate (normal,
CC	cancerous) and colon epithelium. Human prollylhydroxylase-like protein
CC	and polynucleotides and polypeptides, agonists and antagonists can be used in
CC	the prevention, amelioration or correction of dysfunctions or diseases
CC	including cancer, cardiovascular disorders, chronic obstructive pulmonary
CC	disease, athma, genitourinary disorders, and central nervous system
CC	disorders (claimed), e.g. schizophrenia, Alzheimer's disease and
CC	Parkinson's disease
XX	
XX	Sequence 747 AA;
SQ	
Query Match            85.7%;    Score 3316;    DB 7;    Length 747;	
Best Local Similarity    88.2%;    Pred. No. 1.5e-283;	
Matches    623;    Conservative    33;    Mismatches    48;    Indels        2;    Gaps        1;	
QY	25 ESEAGGCVTPDLLFAEGTAAYARGDPGWGLSMERALRSRAALRALRLRCRTQCAADFP 84
DB	25 ESEPGDWVAAPDLLVYAEGTAAYSRRDWPGLNMERALSRAALRALRLRCRTRCATLP 84
QY	85 WELDDPWSSP--AQASGAGALEDLSPFGLLLRACLRLCPGAHHSISEMELEFRK 142
DB	85 WAPDDLGLDPDSLSDPGGAAALHDUFFGAVLRAACLRCLGGPPSAHLSEELDLEFNK 144
QY	143 RSPYNVLQVAFKINKLEKAVAAAHTFFVGNEPEHEMQQNLDYYQTMSGVKAEADFKDLET 202
DB	145 RSPYNVLQVAFKINKLEKAVAAAHTFFVGNEPEHEMVRQNLDYYQTMSGVKAEADFDLEA 204
QY	203 QPHMQEFLRGVRLYSEEQOEAVPHLEALQLQEVFYAVECRALCEGPYDYGNYLEYNA 262
DB	205 KPHMHFRLGVRLYSEKQPQAVPHLEALQLQEVFYADEECRALCEGPYDYGNYLDYSA 264
QY	263 DLFOALTTHVIQVNLCKNCVTETASHPREKPFDFLP SHYNYLQFAYNIGNYTQAGE 322
DB	265 DLFOALTTHVQVNLCKNCVTETASHPREKPFDFLP SHYNYLQFAYNIGNYTQAI 324
QY	323 CAKTYLLFFPNDEVMNQNTLAYAAMLGEBTRTSIGPRESAKVQRSLLEKEILLFFAYDV 382
DB	325 CAKTYLLFFENDEVHQNTLAYVTAMLGEEASGISPRENAEYRRPNLLEKEILLFAYDI 384
QY	383 FGI PFVDPSWTPEEVI PKRLQEKQKSERETA VRI SQEIGNLMKETTELVEEKTESLDV 442
DB	385 FGI PFVDPSWTPEEVI PKRLQEKQKSERETA VRI SQEIGNLMKETTELVEEKTESLDV 444
QY	443 SRLTREGGPLLYEGISLTWN SKLLNGYQVWMDGVISDHCEQELQRLTNVAATSGDG YRG 502
DB	445 SRLTREGGPLLYEGISLTWN SKVLNGYQVWMDGVISDDCEQELQRLTNAATSGDG YRG 504
QY	503 QTS PHTPNKFYGVTVFKALKLGQEKKVPLOS AHLYNNYTEKVRRIMESYFR LDTPLYFS 562
DB	505 QTS PHTPNKFYGVTVFKALKLGQEKKVPLOS ARMYNNYTEKVRRIMESYFR LDTPLYFS 564

QY 563 YSHLVCRFAIEVQAEKDDSHPHVNDNCILNAETLVCKVEPPAYTRDYSAILYLNGDF 622  
 DB 565 YSHFVCRFAIEVQAEKDDSHPHVNDNCILNAEAFMCIEPPAYTRDYSAILYLNGDF 624  
 QY 623 DGGNPFYFTELDKATVTAEVQCCGAVGSSGTENPHGVKAVTRGQCAALWFTLDPRH 682  
 DB 625 DGGNPFYFTELDKATVTAEVQCCGAVGSSGTENPHGVKAVTRGQCAALWFTLDPRH 684  
 QY 683 SERDRVQADDLVKMLFSPPEMDLQEQPLDAQGPPPPAQSLSGS 728  
 DB 685 SERDRVQADDLVKMLFSPPEMDLQEQPLDAQGPPPPAQSLSGS 730

# RESULT 14 AAE04244

ID AAE04244 standard; protein; 592 AA.

AC AAE04244;

DT 09-AUG-2001 (first entry)

DE Human gene 2 encoded secreted protein fragment, SEQ ID NO:100.

XX Human; secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;  
 KW endocrine disorder; infection; wound healing; vulnery; cell culture;  
 KW chemotaxis; food additive; binding partner identification.

XX Homo sapiens.

XX WO200136432-A2.

XX 25-MAY-2001.

XX 15-NOV-2000; 2000WO-US031162.

XX 19-NOV-1999; 99US-0166415P.

XX 30-JUN-2000; 2000US-0215136P.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Komatsoulis GA, Baker KP, Young PE;

DR WPI; 2001-343793/36.

PT Isolated nucleic acid molecule encoding a human secreted protein is used

PT in preventing, treating or ameliorating a medical condition.

PS Disclosure; Page 470-472; 509pp; English.

XX AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted  
 CC protein genes, and AAE04199-AAE04239 represent the proteins they encode.  
 CC AAE04240-AAE04297 represent human secreted protein fragments or variants.  
 CC The secreted proteins and their genes are useful for preventing, treating  
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
 CC Pathological conditions can be diagnosed by determining the amount of the  
 CC new protein in a sample or by determining the presence of mutations in  
 CC the new genes. Specific uses are described for each of the 18 genes,  
 CC based on the tissues in which they are most highly expressed, and include  
 CC developing products for the diagnosis or treatment of proliferative  
 CC disorders, cancer, tumours, foetal and developmental abnormalities,  
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies, disease,  
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,

CC angiogenic disorders, kidney disorders, gastrointestinal disorders,  
 CC pregnancy-related disorders, endocrine disorders, and infections. The  
 CC proteins can also be used to aid wound healing and epithelial cell  
 CC proliferation, to prevent skin ageing due to sunburn, to maintain organs  
 CC before transplantation, for supporting cell culture of primary tissues,  
 CC to regenerate tissues, to identify their cognate ligands or binding  
 CC partners, and in chemotaxis, and can be used as a food additive or  
 CC preservative to modify storage properties. Antibodies specific for a  
 CC protein of the invention can be used in alleviating symptoms associated  
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
 CC present sequence represents a human secreted protein fragment referred to  
 CC in the disclosure of the invention

XX Sequence 592 AA;

Query Match 80.5%; Score 3115; DB 4; Length 592;

Best Local Similarity 99.7%; Pred. No. 6.1e-266;

Matches 590; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 145 PYNLYQAVYFKINKLEKAVAAHTFFVGNPHEMCOQNLDDYQTMGSKYKADFKDLETQP 204

DB 1 PYNLYQAVYFKINKLEKAVAAHTFFVGNPHEMCOQNLDDYQTMGSKYKADFKDLETQP 60

QY 205 HMQEFLGLVRLYSERQPOEAVPHLEAALQEVFVAYEECRALCEGPDYDGYNLYEYNADL 264

DB 61 HMQEFLGLVRLYSERQPOEAVPHLEAALQEVFVAYEECRALCEGPDYDGYNLYEYNADL 120

QY 265 FOAITHYIQLNCKQNCVTELASHPSREKPEPFLPSHYNYLOPAYNIGNYTOAGBCA 324

DB 121 FOAITHYIQLNCKQNCVTELASHPSREKPEPFLPSHYNYLOPAYNIGNYTOAGBCA 180

QY 325 KYTLFFPNDEVMQNLAAYAAMLGEHTRISIGPRESAKYRQSLLEKELLFFAYDVFG 384

DB 181 KYTLFFPNDEVMQNLAAYAAMLGEHTRISIGPRESAKYRQSLLEKELLFFAYDVFG 240

QY 385 IPFVDPDSWTPEEVIKSLQKQKSERETAVRISOEIGNLMKEIETLVEEKTESLDVSR 444

DB 241 IPFVDPDSWTPEEVIKSLQKQKSERETAVRISOEIGNLMKEIETLVEEKTESLDVSR 300

QY 445 LTREGGPLLYEGISLTMNSKLINGQYRVVMDGVSDEHCOELQSLTNVAATSGGYRGQT 504

DB 301 LTREGGPLLYEGISLTMNSKLINGQYRVVMDGVSDEHCOELQSLTNVAATSGGYRGQT 360

QY 505 SPHTPNEKFGYTVTFKALKQEGKVPLOQSAHLYNVTKEVRIMESYFRDLTPLYFSYS 564

DB 361 SPHTPNEKFGYTVTFKALKQEGKVPLOQSAHLYNVTKEVRIMESYFRDLTPLYFSYS 420

QY 565 HLVCRTAIEEVOAERKDDSHPHVNDNCILNAETLVCKVEPPAYTRDYSAILYLNGDFG 624

DB 421 HLVCRTAIEEVOAERKDDSHPHVNDNCILNAETLVCKVEPPAYTRDYSAILYLNGDFG 480

QY 625 GNIFYTELDKATVTAEVQCCGAVGSSGTENPHGVKAVTRGQCAALWFTLDPRHSE 684

DB 481 GNIFYTELDKATVTAEVQCCGAVGSSGTENPHGVKAVTRGQCAALWFTLDPRHSE 540

QY 685 RDRVQADDLVKMLFSPPEMDLSQEQPLDAQGPPPPAQSLSGSSEKPKDEL 736

DB 541 RDRVQADDLVKMLFSPPEMDLSQEQPLDAQGPPPPAQSLSGSSEKPKDEL 592

## RESULT 15

AAE04244

ID AAE04244 standard; protein; 542 AA.

AC AAE04244;

DT 27-FEB-2001 (first entry)

DE Mouse tumour suppressor Gros1-S protein SEQ ID NO:8.

XX Tumour suppressor; Gros1-L; Gros1-S; cell proliferation; regulation;  
 KW cancer; cytostatic; gene therapy.

Job time : 110 secs

```
XX Mus musculus.
XX
XX WO2000065047-A1.
XX
XX 02-NOV-2000.
XX
XX 26-APR-2000; 2000MO-JP002731.
XX
XX 26-APR-1999; 99JP-00119806.
XX
XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
XX Wadhwa R, Sugihara T, Yoshida A;
XX
XX WPI: 2000-687340/67.
XX
XX N-PSDB; AAC64727.
XX
XX Full-length tumor suppressor genes encoding Gros1-L, Gros1-S and mouse
XX PT homologs participating in regulation of cell proliferation, useful in
XX PT development of preventives and remedies of cancer.
XX
XX Claim 1; Page 104-107; 114pp; Japanese.
XX
XX The present sequence represents the mouse tumour suppressor designated
XX CC Gros1-S. Gros1-L and Gros1-S have cytostatic activity and can be used in
XX CC gene therapy. Gros1-L and Gros1-S genes are useful in the development of
XX CC drugs used to treat and prevent cancer
XX
XX Sequence 542 AA;
XX
Query Match          61.8%; Score 2384; DB 3; Length 542;
Best Local Similarity 87.9%; Pred. No. 2.2e-201;
Matches 451; Conservative 25; Mismatches 35; Indels 2; Gaps 1;
Qy 25 ESAGAGMGTPTDLLFAEGTAAYARGDWPVGVLSMERALRSRAALRALRLRCRTQCAADFP 84
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 25 ESFPGHVDVAAPDILYAGTAAYSRDWPVGVILNMERALRSRAALRALRLRCRTCAATLP 84
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 85 WELDPPWSPSP--AQASGAGALRDLPFGGLLRRAACLRCLGPPAAHSLSEMELEPRK 142
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 85 WAPDLDLGPDPSPSQDPGAAALHDLRFFGAVLRRAACLRCLGPPSAHLISELDLEFNK 144
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 143 RSPVNYLOVAYFKINKLEKAVAAAHFFVGNPEHMEVQONLDYYQTMVGKKEADFKDLET 202
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 145 RSPNYLQVAYFKINKLEKAVAAAHFFVGNPEHMEVQONLDYYQTMVGKKEADFKDLEA 204
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 203 QPHMQEFLGLVRLYSEEPQAEVPHLEAALQEFVAYEECRALCEGPDYDGYNYLEYNA 262
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 205 KPHMHEFLGLVRLYSEEPQAEVPHLEAALQEFVAYEECRALCEGPDYDGYNYLDYSA 264
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 263 DLFOAITDHYIOVLNCKQNCVTELASHPSREKPFEDPLPSHNYLQFAYYNYIGNYTOAGE 322
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 265 DLFOAITDHYVQVLNCKQNCVTELASHPSREKPFEDPLPSHNYLQFAYYNYIGNYTOAIE 324
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 323 CAKTYLLFFNDENVNQNLAYAAMLGCEHTRGTPRESAKYRQRSLLKELLFFAYDV 382
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 325 CAKTYLLFFNDENVHQNLAYYAMLGEEASSISPRNAEYRRNLLKELLFFAYDI 384
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 383 FGIPFVDPDSWTPEVTPKQLQKQKSERETAVRISOEIGNLMKEIETLVEEKTESLDV 442
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 385 FGIPFVDPDSWTPEVTPKQLQKQKSERETAVRISOEIGNLMKEIETLVEEKTESLDV 444
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 443 SRLTREGGPIILYEGISLTMSKLYNGYQVVMDSVSDHCEQLQRLTNVAATSGDGYRG 502
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 445 SRLTREGGPIILYEGISLTMSKLYNGYQVVMDSVSDHCEQLQRLTNVAATSGDGYRG 504
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 503 QTSPTTNEKFGYGVTFKALKLQEGKVPLOSA 535
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 505 QTSPTTNEKFGYGVTVLALKLQEGKVPLOSA 537
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: July 18, 2004, 09:40:09

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 09:41:58 ; Search time 62 Seconds  
(without alignments)

3354.112 Million cell updates/sec

Title: US-10-045-815-4

Perfect score: 736

Sequence: 1 MAVRALKLLTLLAVVAAS.....PPEPAQESLSGSESKPKDEL 736

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A Geneseq 29Jan04.\*

- 1: geneseqp1980s.\*
- 2: geneseqp1990s.\*
- 3: geneseqp2000s.\*
- 4: geneseqp2001s.\*
- 5: geneseqp2002s.\*
- 6: geneseqp2003as.\*
- 7: geneseqp2003bs.\*
- 8: geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	736	100.0	736	3 AAB36392	Aab36392 Human tum
2	634	86.1	736	5 AAB69129	Abp69129 Human pol
3	433	58.8	736	4 AAB93125	Aab93125 Human pro
4	433	58.8	736	4 AAB88373	Aab88373 Human mem
5	433	58.8	736	5 AAU09861	Aau09861 Novel hum
6	407	55.3	736	4 AAB93142	Aab93142 Human pro
7	390	53.0	592	4 AAE04244	Aae04244 Human gen
8	380	51.6	806	5 ABG66709	Abg66709 Human nov
9	380	51.6	806	5 ABG66686	Abg66686 Human nov
10	360	48.9	363	3 AAB36391	Aab36391 Human tum
11	281	38.2	804	4 AAB94482	Aab94482 Human pro
12	275	37.4	708	5 AAU09860	Aau09860 Novel hum
13	267	36.3	337	4 AAE04246	Aae04246 Human gen
14	244	33.2	706	5 AABP69128	Abp69128 Human pol
15	219	29.8	359	4 AAE04200	Aae04200 Human gen
16	219	29.8	359	5 ABG64502	Abg64502 Human alb
17	219	29.8	408	5 AABP69127	Abp69127 Human pol
18	188	25.5	400	5 ABG73582	Abg73582 Human leu
19	179	24.3	182	6 ABU70516	Abu70516 Human adi
20	172	23.4	217	3 AAB42734	Aab42734 Human ORF
21	101	13.7	173	4 AAE04217	Aae04217 Human gen
22	101	13.7	173	5 ABG64501	Abg64501 Human alb
23	91	12.4	747	3 AAB36393	Aab36393 Mouse tum
24	91	12.4	747	7 ABR63160	Abri63160 Murine po
25	82	11.1	542	3 AAB36394	Aab36394 Mouse tum

26	72	9.8	81	4 AAE04245	Aae04245 Human gen
27	60	8.2	60	4 AAE04247	Aae04247 Human gen
28	56	7.6	227	4 ABG16396	Abg16396 Novel hum
29	13	1.8	219	4 AAU87297	Aau87297 Novel gen
30	13	1.8	267	4 AAU17171	Aau17171 Novel sig
31	13	1.8	267	4 AAU87588	Aau87588 Novel gen
32	13	1.8	267	7 ADB93879	Adb93879 Human nov
33	13	1.8	282	7 ADB82712	Adb82712 Human pro
34	13	1.8	527	4 AAB93094	Aab93094 Human pro
35	13	1.8	708	6 ADA54992	Ada54992 Human pro
36	13	1.8	708	7 ABR63159	Abri63159 Human pro
37	10	1.4	286	4 AEG16279	Aeg16279 Novel hum
38	9	1.2	15	5 ABG73583	Abg73583 Human leu
39	9	1.2	266	4 AAU32739	Aau32739 Novel hum
40	9	1.2	364	6 ABQ07268	Abq07268 Human p53
41	8	1.1	10	4 AAG95546	Aag95546 Human com
42	8	1.1	47	4 AAM94163	Aam94163 Human rep
43	8	1.1	47	4 ABB95628	Abb95628 Human tes
44	8	1.1	222	6 ABU26511	Abu26511 Protein e
45	8	1.1	264	6 ABU11680	Abu11680 Human MDD

#### ALIGNMENTS

#### RESULT 1

AAB36392  
ID AAB36392 standard; protein; 736 AA.

XX AAB36392;

DT 27-FEB-2001 (first entry)

XX Human tumour suppressor Gros1-S protein SEQ ID NO:4.

KW Tumour suppressor; Gros1-L; Gros1-S; cell proliferation; regulation;  
KW cancer; cytostatic; gene therapy.

XX Homo sapiens.

PN WO200065047-A1.

PD 02-NOV-2000.

PF 26-APR-2000; 2000WO-JP002731.

PR 26-APR-1999; 95JP-00118806.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX Wadhwa R, Sugihara T, Yoshida A;

DR WPI; 2000-687340/67.

DR N-PSDB; AAC64725.

XX Full-length tumor suppressor genes encoding Gros1-L, Gros1-S and mouse  
PT homologs participating in regulation of cell proliferation, useful in  
development of preventives and remedies of cancer.

XX Claim 1; Page 78-83; 114pp; Japanese.

CC The present sequence represents the human tumour suppressor designated  
Gros1-S. Gros1-L and Gros1-S have cytostatic activity and can be used in  
gene therapy. Gros1-L and Gros1-S genes are useful in the development of  
drugs used to treat and prevent cancer

XX Sequence 736 AA;

Query Match 100.0%; Score 736; DB 3; Length 736;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 736; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAVRALKLLTLLAVVAASQAQVESEAGWGVTPDLLFAEGTAAYARGDWPGVLSMER 60

Db	1	MAVFAALKLLITLLAVAAASQAEVSEBAGWMTPTDLLFAEGTAAYARGDWPQVLSMER	60
Qy	61	ALRSRAALRALRLRCRTCCADFPWELDPWSPPACASGAGALRDLSPFGGLLRAACL	120
Db	61	ALRSRAALRALRLRCRTCCADFPWELDPWSPPACASGAGALRDLSPFGGLLRAACL	120
Qy	121	RRCLGPPAAHSLSEMELEFRKSPYNYLOVAFKINKLEKAVAAAHTFFVGNPEMEMQ	180
Db	121	RRCLGPPAAHSLSEMELEFRKSPYNYLOVAFKINKLEKAVAAAHTFFVGNPEMEMQ	180
Qy	181	QNLDDYQTMGSKVKEADFKDLSTQPHMQEFLGVLRLYSEEQPOEAVPHLEALQYFYAYE	240
Db	181	QNLDDYQTMGSKVKEADFKDLSTQPHMQEFLGVLRLYSEEQPOEAVPHLEALQYFYAYE	240
Qy	241	ECRALCEGPDYDGYNYLEYNADLFOAITDHYIQVLNCKNCVTELASHPSREKPPEDFL	300
Db	241	ECRALCEGPDYDGYNYLEYNADLFOAITDHYIQVLNCKNCVTELASHPSREKPPEDFL	300
Qy	301	PSHYNILQFAYNIGNTQAGECAKTYLLFPFNDENVNQNLAAYAAMLGSEHTRSIGPRE	360
Db	301	PSHYNILQFAYNIGNTQAGECAKTYLLFPFNDENVNQNLAAYAAMLGSEHTRSIGPRE	360
Qy	361	SAKEYRORSLEKELLFPAYDFVFGIPFVDPDSWTPPEVIPKRLQEKOKSERETAVRISQE	420
Db	361	SAKEYRORSLEKELLFPAYDFVFGIPFVDPDSWTPPEVIPKRLQEKOKSERETAVRISQE	420
Qy	421	IGNLMKBIETLVEBKTESLDVSLTRREGGPLLVEGISTLWNKLLNGYQRVVMGVID	480
Db	421	IGNLMKBIETLVEBKTESLDVSLTRREGGPLLVEGISTLWNKLLNGYQRVVMGVID	480
Qy	481	HECELORLTNVAATSGDVGRCQTSPTPNEKFGYVTVFKALKLGOEKGKVPLOSAHLYN	540
Db	481	HECELORLTNVAATSGDVGRCQTSPTPNEKFGYVTVFKALKLGOEKGKVPLOSAHLYN	540
Qy	541	VTEKVRIMESYFRDLTPLYFSYSHLVCRVTAIEVQAEKRDQDHPHVDNCILNAETLVC	600
Db	541	VTEKVRIMESYFRDLTPLYFSYSHLVCRVTAIEVQAEKRDQDHPHVDNCILNAETLVC	600
Qy	601	VKEPPAYTFRYSAILYLINGDFDGNFYFTELDAKTVTAEVQPCGKRAVFGSGTENPHG	660
Db	601	VKEPPAYTFRYSAILYLINGDFDGNFYFTELDAKTVTAEVQPCGKRAVFGSGTENPHG	660
Qy	661	VKATRGORCAIALWFTLDRHSERDRVQADDLVKMLFSPPEMDLSQEQPLDAQOQPPPEP	720
Db	661	VKATRGORCAIALWFTLDRHSERDRVQADDLVKMLFSPPEMDLSQEQPLDAQOQPPPEP	720
Qy	721	AOESLSGSESKPKDEL 736	
Db	721	AOESLSGSESKPKDEL 736	
RESULT 2			
ABP69129	ID	ABP69129 standard; protein; 736 AA.	
XX	AC	ABP69129;	
XX	XX		
DT	20-JAN-2003	(first entry)	
XX	DE	Human polypeptide SEQ ID NO 1176.	
KW	KW	Human; genome mapping; gene therapy; food supplement; virus; fungus;	
KW	KW	cell-proliferative disorder; neurodegenerative disease; bacterial;	
KW	KW	Parkinson's disease; Alzheimer's disease; autoimmune disease;	
KW	KW	multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;	
KW	KW	arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;	
KW	KW	antiParkinsonian; antidiabetic; immunosuppressive; dermatological;	
KW	KW	haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;	
OS	XX	Homo sapiens.	
XX	XX		

PN	WC200270539-A2.		
XX	12-SEP-2002.		
XX	05-MAR-2002; 2002WO-US0005095.		
XX	05-MAR-2002; 2001US-00799451.		
XX	(HYSE-) HYSEQ INC.		
PI	Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;		
PI	Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;		
PI	Wehrman T, Wang J, Wang D, Drmanac R;		
XX	WPI; 2002-759812/82.		
DR	N-PSDB; AB211346.		
XX	New polynucleotides comprising sequences assembled from expressed		
PT	sequence tags (ESTs), useful for treating cell-proliferative,		
PT	neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet		
PT	or coagulation disorders.		
XX	Claim 9; SEQ ID NO 1176; 1012bp + Sequence Listing; English.		
XX	The invention relates to an isolated polynucleotide (I) comprising a		
CC	nucleotide sequence selected from any of 948 sequences (AB211119-		
CC	AB212086) or their mature protein coding portion, active domain coding		
CC	protein or complementary sequences. The polynucleotides are useful for		
CC	identifying expressed genes or for physical mapping of human genome. The		
CC	encoded polypeptides (ABP6902-ABP6949) are useful as molecular weight		
CC	markers, as a food supplement, for generating antibodies, in medical		
CC	imaging, screening and diagnostic assays and for treating cell-		
CC	proliferative disorders (cancer), neurodegenerative diseases (Parkinson's		
CC	or Alzheimer's disease), autoimmune diseases (multiple sclerosis,		
CC	diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,		
CC	platelet or coagulation disorders, wound, burns, incision, ulcers, liver		
CC	or lung fibrosis, infections (bacterial, viral, fungal, parasitic),		
CC	arthritis, etc. Note: The sequence data for this patent did not form part		
CC	of the printed specification, but was obtained in electronic format		
CC	directly from WIPO at ffp.wipo.int/pub/published_pct_sequences		
XX	Sequence 736 AA;		
SQ			
Query Match 86.1%; Score 634; DB 5; Length 736;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 634; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	103	ALRDLSPFGGLLRRACLRCLGPPAAHSLSEMELEFRKSPYNYLOVAFKINKLEKA	162
Db	103	ALRDLSPFGGLLRRACLRCLGPPAAHSLSEMELEFRKSPYNYLOVAFKINKLEKA	162
Qy	163	VAAAHTEFFVGNPEHEMOMQNDYYQTMGSKVKEADFKDLSTQPHMQEFLGVLRLYSEEQPQ	222
Db	163	VAAAHTEFFVGNPEHEMOMQNDYYQTMGSKVKEADFKDLSTQPHMQEFLGVLRLYSEEQPQ	222
Qy	223	EAVPHLEALQYFYVAYEECRALCEGPDYDGYNYLEYNADLFOAITDHYIQVLNCKNC	282
Db	223	EAVPHLEALQYFYVAYEECRALCEGPDYDGYNYLEYNADLFOAITDHYIQVLNCKNC	282
Qy	283	VTELASHPSREKPPEDFLPSHYNILQFAYNIGNTQAGECAKTYLLFPFNDENVNQNLA	342
Db	283	VTELASHPSREKPPEDFLPSHYNILQFAYNIGNTQAGECAKTYLLFPFNDENVNQNLA	342
Qy	343	YYAAMLGSEHTRSIGPRESAKYRORSLEKELLFPAYDFVFGIPFVDPDSWTPPEVIPKR	402
Db	343	YYAAMLGSEHTRSIGPRESAKYRORSLEKELLFPAYDFVFGIPFVDPDSWTPPEVIPKR	402
Qy	403	LOEKOKSERETAVRISQIEGNLMKBIETLVEBKTESLDVSLTRREGGPLLVEGISTLWN	462
Db	403	LOEKOKSERETAVRISQIEGNLMKBIETLVEBKTESLDVSLTRREGGPLLVEGISTLWN	462
Qy	463	SKLLNGYQRVVMGVIDSDHECOELRLTNVAATSGDVGRCQTSPTPNEKFGYVTVFKAL	522

463 SKLLNGYQVRVMDGVISDHECQELQRLTNVAATSGDYGRTSPHTPNEKFGVTVFKAL 522  
523 KLGQEGKVPLOSALHYNNVTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDD 582  
523 KLGQEGKVPLOSALHYNNVTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDD 582  
583 SHPVHVDNCLINAEITLVCKVEKPPAYTFRDYSAILYLNGDFDGNFVFTTDLAKTVTAEVQ 642  
583 SHPVHVDNCLINAEITLVCKVEKPPAYTFRDYSAILYLNGDFDGNFVFTTDLAKTVTAEVQ 642  
643 PQCGRAVSGSGTENPHGVKAVTRGQRCALMFTLDPHRSERDRVQADLVKMLFSPPE 702  
643 PQCGRAVSGSGTENPHGVKAVTRGQRCALMFTLDPHRSERDRVQADLVKMLFSPPE 702  
703 MDLSQQLDQAQQPPPEPAQESLSGSESKPKDEL 736  
703 MDLSQQLDQAQQPPPEPAQESLSGSESKPKDEL 736

RESULT 3  
AAB93215  
ID AAB93215 standard; protein; 736 AA.

XX AAB93215;  
XX 26-JUN-2001 (first entry)  
XX Human protein sequence SEQ ID NO:12194.  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX Homo sapiens.  
XX EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-00116126.  
XX 29-JUL-1999; 99JP-00248036.  
XX 27-AUG-1999; 99JP-00300253.  
XX 11-JAN-2000; 2000JP-00118776.  
XX 02-MAY-2000; 2000JP-00183767.  
XX 09-JUN-2000; 2000JP-00241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID NO 12194; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAB93166 to AAB9328 and AAB9329 to AAB9344 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAB93629 to AAB93632 represent oligonucleotides, all of which are used in the exemplification of the present invention  
XX Sequence 736 AA;  
SQ

Query Match 58.8%; Score 433; DB 4; Length 736;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAVRALKLLTLLAVAAASQAEVSEAGWGMVTPDLLFAEGTAAYARGDMPGVVLSNER 60  
DB 1 MAVRALKLLTLLAVAAASQAEVSEAGWGMVTPDLLFAEGTAAYARGDMPGVVLSNER 60  
QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPSPAQASGAALRLDLSFFGGLLRAACL 120  
DB 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPSPAQASGAALRLDLSFFGGLLRAACL 120  
QY 121 RRCIGPPAAHSISEMELEFRKSPYNYLVAYKINKLEKAVAAHTFFVGNDEHMEMQ 180  
DB 121 RRCIGPPAAHSISEMELEFRKSPYNYLVAYKINKLEKAVAAHTFFVGNDEHMEMQ 180  
QY 181 QNLDDYQTMGSKVADFKDLETPHMQEFRGLGVLYSEEQPQEAHPHLEAALQYFYVAYE 240  
DB 181 QNLDDYQTMGSKVADFKDLETPHMQEFRGLGVLYSEEQPQEAHPHLEAALQYFYVAYE 240  
QY 241 ECRALCEGPDYDGYNYLEYNADLFOAITDHYIQLNCKQNCVTELASHPSREKPFDFL 300  
DB 241 ECRALCEGPDYDGYNYLEYNADLFOAITDHYIQLNCKQNCVTELASHPSREKPFDFL 300  
QY 301 PSHYNYLOFAYNYIGNYTQAGECAKTYLLFPFNDVNMNLAYYAAMLGEHTSISGRE 360  
DB 301 PSHYNYLOFAYNYIGNYTQAGECAKTYLLFPFNDVNMNLAYYAAMLGEHTSISGRE 360  
QY 361 SAKYRQSRSLLEKELLFPAYDVFGIPFVDPDSWTPPEVPIKRLQEKQKSERETAVRISQE 420  
DB 361 SAKYRQSRSLLEKELLFPAYDVFGIPFVDPDSWTPPEVPIKRLQEKQKSERETAVRISQE 420  
QY 421 IGNLKKEITLVEKTKESLDVSRITREGGPLYEGISLTWNKLANGYQVRVMDGVISD 480  
DB 421 IGNLKKEITLVEKTKESLDVSRITREGGPLYEGISLTWNKLANGYQVRVMDGVISD 480  
QY 481 HECQELQRLTNVAATSGDYGRTSPHTPNEKFGVTVFKALKLGQEGKVPLOSALHYNN 540  
DB 481 HECQELQRLTNVAATSGDYGRTSPHTPNEKFGVTVFKALKLGQEGKVPLOSALHYNN 540  
QY 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPVHVDNCLINAEITLV 600  
DB 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKDDSHPVHVDNCLINAEITLV 600  
QY 601 VKEPPAYTFRDYSAILYLNGDFDGNFVFTTDLAKTVTAEVQPCGQRAVGSSTENPHG 660  
DB 601 VKEPPAYTFRDYSAILYLNGDFDGNFVFTTDLAKTVTAEVQPCGQRAVGSSTENPHG 660  
QY 661 VKAVTRGQRCALMFTLDPHRSERDRVQADLVKMLFSPPEMDLSQQLDQAQQPPPEP 720  
DB 661 VKAVTRGQRCALMFTLDPHRSERDRVQADLVKMLFSPPEMDLSQQLDQAQQPPPEP 720  
QY 721 AQESLSGSESKPKDEL 736  
DB 721 AQESLSGSESKPKDEL 736

RESULT 4  
AAB88373  
ID AAB88373 standard; protein; 736 AA.  
XX AAB88373;  
AC AAB88373;

XX DT 23-MAY-2001 (first entry)  
 XX DE Human membrane or secretory protein clone PSEC0109.  
 XX KW Human; secretory protein; membrane protein; vaccine; gene therapy;  
 KW rheumatoid arthritis; diabetes.  
 XX OS Homo sapiens.  
 XX FN EP1067182-A2.  
 XX PD 10-JAN-2001.  
 XX PF 07-JUL-2000; 2000EP-00114090.  
 XX PR 08-JUL-1999; 99JP-00194179.  
 PR 11-JAN-2000; 2000JP-00118775.  
 PR 02-MAY-2000; 2000JP-00183766.  
 XX (HELI-) HELIX RES INST.  
 PA Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;  
 PI WPI; 2001-093989/11.  
 DR N-PSDB; AAF93800.  
 XX Nucleic acids encoding secretory proteins/membrane proteins, useful in  
 PT gene therapy or as candidate target molecules in drug development.  
 XX Claim 1; SEQ ID NO 114; 609pp + Sequence Listing; English.  
 XX This invention relates to nucleic acid sequences AAF93744 - AAF93916  
 CC which encode human secretory or membrane proteins represented by AAB88317  
 CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and  
 CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the  
 CC invention. The invention also includes methods for the production of  
 CC antibodies directed against the proteins, and cDNA sequences, which can  
 CC be used in vaccines. The polynucleotide sequences can be used in gene  
 CC therapy. The polynucleotide sequences and the proteins they encode may be  
 CC used in the prevention, treatment and diagnosis of diseases associated  
 CC with inappropriate secretory protein/membrane protein expression. The  
 CC nucleic acids and complementary sequences may also be used as DNA probes  
 CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect  
 CC and quantitate the presence of similar nucleic acid sequences in samples.  
 CC They may also be used to study the expression and function of secretory  
 CC proteins/membrane polypeptides and their role in metabolism. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC against them and in assays to identify modulators (agonists and  
 CC antagonists) of expression and activity. The antibodies and antagonists  
 CC may also be used as therapeutic agents to down regulate expression and  
 CC activity. The antibodies may also be used as diagnostic agents for  
 CC detecting the presence of the polypeptides in samples (e.g. by enzyme  
 CC linked immunosorbant assay (ELISA)). Examples of diseases which may be  
 CC treated include rheumatoid arthritis and diabetes

Query Match 58.8%; Score 433; DB 4; Length 736;  
 Best Local Similarity 99.6%; Pred. No. 0;  
 Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MAVFALKLLTLLAVAAASQAEVESEAGMGWTPDLLFAEGTAAYARGDWPQVLSMER 60  
 Db 1 MAVFALKLLTLLAVAAASQAEVESEAGMGWTPDLLFAEGTAAYARGDWPQVLSMER 60  
 Qy 61 ALRSEALRALRLRCRTQCAADFPWELDPDWSPSQAASGAGALRDLSPFGGLLRAACL 120  
 Db 61 ALRSEALRALRLRCRTQCAADFPWELDPDWSPSQAASGAGALRDLSPFGGLLRAACL 120  
 Qy 121 RRCILGPPAAHLSSEMELEFRKRSPPYNYLOVAFKINKLEKAVAAAHFTFVGNPEHMEQ 180  
 Db 121 RRCILGPPAAHLSSEMELEFRKRSPPYNYLOVAFKINKLEKAVAAAHFTFVGNPEHMEQ 180

QY 181 QNLDDYQTMGSKVEADFKDLETQPHMQEFLGVLRYSEEQPQEAHPHLEALQYFVAYE 240  
 DB 181 QNLDDYQTMGSKVEADFKDLETQPHMQEFLGVLRYSEEQPQEAHPHLEALQYFVAYE 240  
 QY 241 ECRALCEGPDYDGYNYLEYNADLFOAITDHYIQLVNLCKQNCVTELASHSPSREKPFDFL 300  
 DB 241 ECRALCEGPDYDGYNYLEYNADLFOAITDHYIQLVNLCKQNCVTELASHSPSREKPFDFL 300  
 QY 301 PSHYNYLOFAYYNYGNITQAGCAKTYLLFPNDEVNMNQLAYYAAMLGEEHTRSIGPRE 360  
 DB 301 PSHYNYLOFAYYNYGNITQAGCAKTYLLFPNDEVNMNQLAYYAAMLGEEHTRSIGPRE 360  
 QY 361 SAKYRQESLLEKELLFPAYDVFGIPFVDPDSWTPEEVIKRLQEKOKSERETAVRISQE 420  
 DB 361 SAKYRQESLLEKELLFPAYDVFGIPFVDPDSWTPEEVIKRLQEKOKSERETAVRISQE 420  
 QY 421 IGNLMKEIETLVEEKTESLDSRLTREGGPLLYEGISLTWNKLLNGYQRVNMDGVISD 480  
 DB 421 IGNLMKEIETLVEEKTESLDSRLTREGGPLLYEGISLTWNKLLNGYQRVNMDGVISD 480  
 QY 481 HECQELQRLTNVAATSGDGYRGQTSPTPNKFFYGVTVFKALKGQSKVPLQSAHLYYN 540  
 DB 481 HECQELQRLTNVAATSGDGYRGQTSPTPNKFFYGVTVFKALKGQSKVPLQSAHLYYN 540  
 QY 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRCTAIEEVQAEKDDSHPVHVDNCILNAETLVC 600  
 DB 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRCTAIEEVQAEKDDSHPVHVDNCILNAETLVC 600  
 QY 601 VKEPAYTFRDYSAILYLNGFDGNGFYFTLDAKTVAEVQPCGAVGSSSTENPHG 660  
 DB 601 VKEPAYTFRDYSAILYLNGFDGNGFYFTLDAKTVAEVQPCGAVGSSSTENPHG 660  
 QY 661 VKAVTRGQRCALALWFLDPRHSERDRVQADDLVKMLFSPPEMDLSOEPLDAQGGPPEP 720  
 DB 661 VKAVTRGQRCALALWFLDPRHSERDRVQADDLVKMLFSPPEMDLSOEPLDAQGGPPEP 720  
 QY 721 AQESLSGSESKPKDEL 736  
 DB 721 AQESLSGSESKPKDEL 736

RESULT 5  
 AAU09861  
 ID AAU09861 standard; protein; 736 AA.  
 XX  
 AC AAU09861;  
 XX  
 DT 26-FEB-2002 (first entry)  
 XX  
 DE Novel human secreted protein #2.  
 XX  
 KW Secreted protein; cytostatic; immunosuppressive; vulnary; vaccine;  
 KW antinflammatory; neuroprotective; nephrotropic; cardiovascular; human;  
 KW cancer; autoimmune disease; wound healing disorder; infection;  
 KW haematopoietic disorder; inflammatory disorder; infertility;  
 KW neurological disease; psychiatric disease; cardiovascular disease;  
 KW respiratory disease; renal; gastrointestinal.

XX Homo sapiens.  
 OS  
 XX WO200179454-A1.  
 PN  
 XX 25-OCT-2001.  
 PD  
 PF 11-APR-2001; 2001WO-US011797.  
 XX  
 PR 13-APR-2000; 2000US-0196603P.  
 PR 24-APR-2000; 2000US-0199417P.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX

PI Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z;  
 XX WPI; 2002-061975/08.  
 DR N-PSDB; AAS17573.  
 XX  
 PT New secreted proteins or polypeptides, useful for treating e.g. cancer,  
 PT autoimmune diseases, wound healing disorder, infections, hematopoietic  
 PT disorders, inflammatory disorders, infertility, cancer.  
 XX  
 PS Claim 1; Page 54-56; 92pp; English.  
 XX  
 CC The invention relates to an isolated novel secreted polypeptide (I) and  
 CC polynucleotide (II). (I) and (II) are useful for treating cancer,  
 CC autoimmune diseases, wound healing disorder, infections, hematopoietic  
 CC disorders, inflammatory disorders, infertility, neurological and  
 CC psychiatric diseases, cardiovascular diseases, respiratory diseases,  
 CC renal diseases, or gastrointestinal diseases. These may also be used to  
 CC treat diseases, abnormalities and disorders caused by abnormal  
 CC expression, production, function and/or metabolism of the genes, as  
 CC vaccines for inducing immunological response in a mammal, and in  
 CC screening methods for detecting the effect of added compounds on the  
 CC production of mRNA and polypeptide in cells. The polypeptides can be used  
 CC as immunogens to produce antibodies immunospecific for the polypeptides,  
 CC and to identify membrane-bound or soluble receptors. The polynucleotides  
 CC may be used as diagnostic reagents, in chromosome localisation studies,  
 CC and in tissue expression studies. The present sequence represents the  
 CC amino acid sequence of novel human secreted protein #2  
 XX  
 SQ Sequence 736 AA;  
 Query Match 58.8%; Score 433; DB 5; Length 736;  
 Best Local Similarity 99.6%; Pred. No. 0;  
 Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 MAVRALKLLTLLAVAAASQAQAEVSEAGAGMTPTDLLFAEGTAAYARGDWPGLVLSMER 60  
 DB 1 MAVRALKLLTLLAVAAASQAQAEVSEAGAGMTPTDLLFAEGTAAYARGDWPGLVLSMER 60  
 QY 61 ALRSRAALRALRLCRTOCAADFPWELDPDWSPSQAQAGALRDLSPFGGLLRAACL 120  
 DB 61 ALRSRAALRALRLCRTOCAADFPWELDPDWSPSQAQAGALRDLSPFGGLLRAACL 120  
 QY 121 RRCGLGPAASLSSEMELEFRKSPYNYLOVAFKINKLEKAVAAAHFTFVGNPEHMEQ 180  
 DB 121 RRCGLGPAASLSSEMELEFRKSPYNYLOVAFKINKLEKAVAAAHFTFVGNPEHMEQ 180  
 QY 181 QNLDDYQTMGSKVEADFKDLETPHMQEFLGLVRLYSEEQPQAVPHLEALQEFYVAYE 240  
 DB 181 QNLDDYQTMGSKVEADFKDLETPHMQEFLGLVRLYSEEQPQAVPHLEALQEFYVAYE 240  
 QY 241 ECRALCEGPDYDGYNYLEYNADLFQAITDHYIQVLNCKQNCVTELASHSPREKPFEDFL 300  
 DB 241 ECRALCEGPDYDGYNYLEYNADLFQAITDHYIQVLNCKQNCVTELASHSPREKPFEDFL 300  
 QY 301 PSHVNYLQFAYNIGNYTOAGECAKTVLLFPFNDVNMQLAYYAAMLGEHTRSIGPRE 360  
 DB 301 PSHVNYLQFAYNIGNYTOAGECAKTVLLFPFNDVNMQLAYYAAMLGEHTRSIGPRE 360  
 QY 361 SAKERYORSILLEKLLFPAYDVGIPVDPSMTPEEVIKRLQEKQKSRERAVRISQE 420  
 DB 361 SAKERYORSILLEKLLFPAYDVGIPVDPSMTPEEVIKRLQEKQKSRERAVRISQE 420  
 QY 421 IGNLMKEIETLVEBKTESLDVSELTEGGLLYEGISLTWNSKLLNGYORVMDGVISD 480  
 DB 421 IGNLMKEIETLVEBKTESLDVSELTEGGLLYEGISLTWNSKLLNGYORVMDGVISD 480  
 QY 481 HECQELQRLTNVAATSGDVGRTQSPHTPNEKFYGVTVTFKALIKLQEGKVPLOSAHLYYN 540  
 DB 481 HECQELQRLTNVAATSGDVGRTQSPHTPNEKFYGVTVTFKALIKLQEGKVPLOSAHLYYN 540  
 QY 541 VTEKVRIMESYFRLDTPFLFYSYHLVCRTAIEVQAEKDDSHPHVNDNCILNAETLVC 600  
 DB 541 VTEKVRIMESYFRLDTPFLFYSYHLVCRTAIEVQAEKDDSHPHVNDNCILNAETLVC 600

QY 601 VKEPPAYTFRDYSAILYLNGDFDGGNFYFTELDAKTVTAEVQPCQGRAVGFSSGTENPHG 660  
 DB 601 VKEPPAYTFRDYSAILYLNGDFDGGNFYFTELDAKTVTAEVQPCQGRAVGFSSGTENPHG 660  
 QY 661 VKAVTRGQRCAIALWFTLDPHRSERDRVQADLVXVLFSPPEMDLSQBPDLAQGGPPEP 720  
 DB 661 VKAVTRGQRCAIALWFTLDPHRSERDRVQADLVXVLFSPPEMDLSQBPDLAQGGPPEP 720  
 QY 721 AQESLSGSESKEPKDEL 736  
 DB 721 AQESLSGSESKEPKDEL 736  
 RESULT 6  
 AAB93142  
 ID AAB93142 standard; protein; 736 AA.  
 AC AAB93142;  
 XX  
 XX 26-JUN-2001 (first entry)  
 DT  
 XX Human protein sequence SEQ ID NO:12045.  
 DE  
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 KW  
 XX Homo sapiens.  
 OS  
 XX EP1074617-A2.  
 PN  
 XX 07-FEB-2001.  
 PD  
 XX 28-JUL-2000; 2000EP-00116126.  
 PF  
 XX 29-JUL-1999; 99JP-00248036.  
 PR  
 XX 27-AUG-1999; 99JP-00300253.  
 PR  
 XX 11-JAN-2000; 2000JP-00118776.  
 PR  
 XX 02-MAY-2000; 2000JP-00183767.  
 PR  
 XX 09-JUN-2000; 2000JP-00241899.  
 PR  
 XX (HELI-) HELIX RES INST.  
 PA  
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX WPI; 2001-318749/34.  
 DR  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 PT length cDNAs defined in the specification, and for the detection and/or  
 PT diagnosis of the abnormality of the proteins encoded by the full-length  
 PT cDNAs.  
 PS Claim 8; SEQ ID NO 12045; 2537pp + Sequence Listing; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602 full-  
 CC length cDNAs defined in the specification. Where a primer set comprises:  
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893

CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention  
XX Sequence 736 AA;

Query Match 55.3%; Score 407; DB 4; Length 736;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 707; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAVRALKLLTLLAVVAASQAQVESEAGWGMVTPDLLFABGTAAAYARGDWPVGVLSMER 60  
DB 1 MAVRALKLLTLLAVVAASQAQVESEAGWGMVTPDLLFABGTAAAYARGDWPVGVLSMER 60

QY 61 ALRGAALRALRLRCRTCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
DB 61 ALRGAALRALRLRCRTCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120

QY 121 RRCGLPPAAHLSBEMEELFRKSPYNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180  
DB 121 RRCGLPPAAHLSBEMEELFRKSPYNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180

QY 181 QNLDYYQTMGSKGKADFQDLTQPHMQBFRGLGVRLYSEEQPOEAVPHLEALQYFYVAYE 240  
DB 181 QNLDYYQTMGSKGKADFQDLTQPHMQBFRGLGVRLYSEEQPOEAVPHLEALQYFYVAYE 240

QY 241 ECRALCEGPDYDGYNYLEYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPFDFL 300  
DB 241 ECRALCEGPDYDGYNYLEYNADLFOAITDHYIQVLNCKQNCVTELASHPSREKPFDFL 300

QY 301 PSHNYLQFAYNYNGYTOGECATYLLFPNDENVNQNLAAYAAMLGBEHTRSIGPRE 360  
DB 301 PSHNYLQFAYNYNGYTOGECATYLLFPNDENVNQNLAAYAAMLGBEHTRSIGPRE 360

QY 361 SAKVROSLLEKELLFPAYDVFGIPFVDPDSWTPPEVIPKRLQEKOKSERETAVRISQE 420  
DB 361 SAKVROSLLEKELLFPAYDVFGIPFVDPDSWTPPEVIPKRLQEKOKSERETAVRISQE 420

QY 421 IGNLMKEITLVEKTESLDVSLRTREGGPLYEGISLTNWSKLLNGYQRYVMDGVISD 480  
DB 421 IGNLMKEITLVEKTESLDVSLRTREGGPLYEGISLTNWSKLLNGYQRYVMDGVISD 480

QY 481 HECQELQRLTNVAATSGDYGEGTSPHTPNEKFGVTVFVKALKGQSKVPLQSAHLIYN 540  
DB 481 HECQELQRLTNVAATSGDYGEGTSPHTPNEKFGVTVFVKALKGQSKVPLQSAHLIYN 540

QY 541 VTEKVRIMESYFRILDTPLYSYSHLVCRTAIEBQVQAEKDDSHPVHVDNCILNAETLVC 600  
DB 541 VTEKVRIMESYFRILDTPLYSYSHLVCRTAIEBQVQAEKDDSHPVHVDNCILNAETLVC 600

QY 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTELDAKTYTAEVQPCGKAVGFSSTENPHG 660  
DB 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTELDAKTYTAEVQPCGKAVGFSSTENPHG 660

QY 661 VKAVTRGORCAIALFWTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEQP 710  
DB 661 VKAVTRGORCAIALFWTLDPHRSERDRVQADDLVKMLFSPPEMDLSQEQP 710

RESULT 7

AAE04244

ID AAE04244 standard; protein; 592 AA.

XX AC AAE04244;

XX AC AAE04244;

XX DT 09-AUG-2001 (first entry)

XX DE Human gene 2 encoded secreted protein fragment, SEQ ID NO:100.

XX DE Human;

XX KW secreted protein; proliferative disorder; cancer; tumour;

XX KW fetal abnormality; developmental abnormality; haematopoietic disorder;

XX KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;

XX KW inflammation; allergy; neurological disorder; Alzheimer's disease;

KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KW skin disorder; psoriasis; sepsis; diabetes; diabetes; atherosclerosis;  
KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;  
KW endocrine disorder; infection; wound healing; vulnery; cell culture;  
KW chemotaxis; food additive; binding partner identification.

OS Homo sapiens.

XX WO200136432-A2.

XX 25-MAY-2001.

XX 15-NOV-2000; 2000WO-US01162.

XX 19-NOV-1999; 98US-016641SP.

XX 30-JUN-2000; 2000US-0215138P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Komatsoulis GA, Baker KP, Young PE;

XX WPI; 2001-343793/36.

XX Isolated nucleic acid molecule encoding a human secreted protein is used

XX in preventing, treating or ameliorating a medical condition.

XX Disclosure; Page 470-472; 509pp; English.

XX AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted

XX protein genes, and AAE04199-AAE04239 represent the proteins they encode.

XX AAE04240-AAE04297 represent human secreted protein fragments or variants.

XX The secreted proteins and their genes are useful for preventing, treating

XX or ameliorating medical conditions, e.g., by protein or gene therapy.

XX Pathological conditions can be diagnosed by determining the amount of the

XX new protein in a sample or by determining the presence of mutations in

XX based on the tissues in which they are most highly expressed, and include

XX developing products for the diagnosis or treatment of proliferative

XX disorders, cancer, tumours, foetal and developmental abnormalities,

XX haematopoietic disorders, diseases of the immune system, AIDS, autoimmune

XX diseases (e.g., rheumatoid arthritis), inflammation, allergies,

XX neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),

XX cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,

XX psoriasis), sepsis, diabetes, atherosclerosis, gastrointestinal disorders,

XX angiogenic disorders, kidney disorders, endocrine disorders, and infections.

XX pregnancy-related disorders, endocrine disorders, and infections. The

XX proteins can also be used to aid wound healing and epithelial cell

XX proliferation, to prevent skin ageing due to sunburn, to maintain organs

XX before transplantation, for supporting cell culture of primary tissues,

XX to regenerate tissues, to identify their cognate ligands or binding

XX partners, and in chemotaxis, and can be used as a food additive or

XX preservative to modify storage properties. Antibodies specific for a

XX protein of the invention can be used in alleviating symptoms associated

XX with the disorders mentioned above, and in diagnostic immunoassays e.g.,

XX radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The

XX present sequence represents a human secreted protein fragment referred to

XX in the disclosure of the invention

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

QY 265 FOAITHYIOVLNCKQNCVTELASHPREKPFDFLPESHYNLYQFAYYNGYNTQAGECA 324  
DB 121 FOAITHYIOVLNCKQNCVTELASHPREKPFDFLPESHYNLYQFAYYNGYNTQAVECA 180  
QY 325 KTYLLFFPNDEVMNQNLAYAAAMLGEEHTRSIGPRESAKYRQSRLLLEKLLFFAYDVFG 384  
DB 181 KTYLLFFPNDEVMNQNLAYAAAMLGEEHTRSIGPRESAKYRQSRLLLEKLLFFAYDVFG 240  
QY 385 IPFVDPDSWTPPEVTPKRLQEKOKSERETAVRISOEIGNLMKEIETLVESKTESLDVSR 444  
DB 241 IPFVDPDSWTPPEVTPKRLQEKOKSERETAVRISOEIGNLMKEIETLVESKTESLDVSR 300  
QY 445 LTRGGGLLYEGISLTWNCKLLNGYQVRVMDGVISDHCEQELQRLTNVAATSGGVRGQT 504  
DB 301 LTRGGGLLYEGISLTWNCKLLNGYQVRVMDGVISDHCEQELQRLTNVAATSGGVRGQT 360  
QY 505 SPHTPNKFKYGVTVFKALKIGQEKVPLQSAHLYNNTKVRIMESYFRLDTPLYESYS 564  
DB 361 SPHTPNKFKYGVTVFKALKIGQEKVPLQSAHLYNNTKVRIMESYFRLDTPLYESYS 420  
QY 565 HLVCRTAEVQAEKDDSHPHVNDNCILNAETLVCVKEPPAYTFRDYSAILYNGDFDG 624  
DB 421 HLVCRTAEVQAEKDDSHPHVNDNCILNAETLVCVKEPPAYTFRDYSAILYNGDFDG 480  
QY 625 GNFTFTELDAKTVAEVQPCGRVAVGSSGTENPHGVKAVTRGQRCALWFTLDPHSE 684  
DB 481 GNFTFTELDAKTVAEVQPCGRVAVGSSGTENPHGVKAVTRGQRCALWFTLDPHSE 540  
QY 685 RDRVQADDLVKMLFSPREMDLSOEPLDAQQGPPPEAQESLSGSESKPKDEL 736  
DB 541 RDRVQADDLVKMLFSPREMDLSOEPLDAQQGPPPEAQESLSGSESKPKDEL 592

RESULT 8  
ABG66709  
ID ABG66709 standard; protein; 806 AA.

AC ABG66709;

DT 30-AUG-2002 (first entry)

XX Human novel polypeptide #44.

XX Human; inflammatory condition; shock; sepsis; immune response; cancer;  
XX wound healing; central nervous system disease; haematopoiesis;  
XX peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;  
XX myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;  
XX cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;  
XX bone degenerative disorder; periodontal disease; reperfusion injury;  
XX lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;  
XX allergic condition; thrombolytic; thrombosis; coagulation disorder;  
XX fungal infection.

XX Homo sapiens.

OS WO200244340-A2.

PN 06-JUN-2002.

XX 30-NOV-2001; 2001WO-US047004.

PF 30-NOV-2000; 2000US-00728952.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;

PI Yanazaki V, Ujwal ML, Drmanac RT;

XX WPI; 2002-508509/54.

DR N-PSDB; ABK94933.

XX Novel nucleic acids and polypeptides for diagnosis, treatment of

PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell

PT disorders, cancer and promoting wound healing.

XX Claim 10; Page 616-617; 672pp; English.

XX The invention relates to human novel polynucleotides and associated  
XX polypeptides. The polynucleotides and polypeptides are useful for  
XX treating inflammatory conditions such as arthritis, nephritis, Crohn's  
XX disease, ischaemia-reperfusion injury, shock, sepsis, immune responses  
XX and cancer and for promoting wound healing. The sequences are used to  
XX induce the proliferation of neural cells and regeneration of nerve and  
XX brain tissue, and are useful for the treatment of central and peripheral  
XX nervous system diseases and neuropathies, such as Alzheimer's disease,  
XX Parkinson's disease, Huntington's disease and amyotrophic lateral  
XX sclerosis. The sequences are involved in chemotactic or chemokinetic  
XX activity, regulation of haematopoiesis, treatment of myeloid or lymphoid  
XX cell disorders and platelet disorders such as thrombocytopenia,  
XX regeneration of bone, cartilage, tendon, ligament and/or nerve tissue  
XX growth, tissue repair, healing of burns, incisions, ulcers, treatment of  
XX osteoporosis, osteoarthritis, bone degenerative disorders and periodontal  
XX disease. The sequences of the invention are also useful for gut  
XX protection or regeneration and treatment of lung or liver fibrosis,  
XX including severe combined immunodeficiency (SCID), bacterial or fungal  
XX infections, autoimmune disorders e.g. multiple sclerosis and myasthenia  
XX gravis, allergic conditions such as asthma, thrombolytic or thrombotic  
XX and coagulation disorders. Sequences ABG66666-ABG66758 represent human  
XX novel polypeptides of the invention

SQ Sequence 806 AA;

Query Match 51.6%; Score 380; DB 5; Length 806;

Best Local Similarity 99.7%; Pred. No. 0;

Matches 580; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 155 KINKLEKAVAAHTFFVGNPEHMEMQONLDYQTMGSKVEADFKDLETPHMQEERLQVR 214

DB 225 KINKLEKAVAAHTFFVGNPEHMEMQONLDYQTMGSKVEADFKDLETPHMQEERLQVR 284

QY 215 LYSSEQPOQAVPHLEAALQEQYFVAYEECRALCEGPDYDGVNLYEYNADLFOAITHYIQ 274

DB 285 LYSSEQPOQAVPHLEAALQEQYFVAYEECRALCEGPDYDGVNLYEYNADLFOAITHYIQ 344

QY 275 VLNCQNCVTELASHPREKPFDFLPESHYNLYQFAYYNGYNTQAGECAKTYLLFFPND 334

DB 345 VLNCQNCVTELASHPREKPFDFLPESHYNLYQFAYYNGYNTQAGECAKTYLLFFPND 404

QY 335 EVMNQNLAYAAAMLGEEHTRSIGPRESAKYRQSRLLLEKLLFFAYDVFGIPFVDPDSWT 394

DB 405 EVMNQNLAYAAAMLGEEHTRSIGPRESAKYRQSRLLLEKLLFFAYDVFGIPFVDPDSWT 464

QY 395 PEEVTPKRLQEKOKSERETAVRISOEIGNLMKEIETLVESKTESLDVSRLLTREGGPLLY 454

DB 465 PEEVTPKRLQEKOKSERETAVRISOEIGNLMKEIETLVESKTESLDVSRLLTREGGPLLY 524

QY 455 EGISLTMSKLLNGYQVRVMDGVISDHCEQELQRLTNVAATSGDGYRGQTSPTTNEKTY 514

DB 525 EGISLTMSKLLNGYQVRVMDGVISDHCEQELQRLTNVAATSGDGYRGQTSPTTNEKTY 584

QY 515 GVTVPKALKIGQEKVPLQSAHLYNNTKVRIMESYFRLDTPLYFSYSHLVCTAIE 574

DB 585 GVTVPKALKIGQEKVPLQSAHLYNNTKVRIMESYFRLDTPLYFSYSHLVCTAIE 644

QY 575 VQAEKDDSHPHVNDNCILNAETLVCVKEPPAYTFRDYSAILYNGDFDGGNFYETELDA 634

DB 645 VQAEKDDSHPHVNDNCILNAETLVCVKEPPAYTFRDYSAILYNGDFDGGNFYETELDA 704

QY 635 KTVTAEVQPCGRVAVGSSGTENPHGVKAVTRGQRCALWFTLDPHSEDRVQADDLV 694

DB 705 KTVTAEVQPCGRVAVGSSGTENPHGVKAVTRGQRCALWFTLDPHSEDRVQADDLV 764

QY 695 KMLFSPREMDLSOEPLDAQQGPPPEAQESLSGSESKPKDEL 736

DB 765 KMLFSPREMDLSOEPLDAQQGPPPEAQESLSGSESKPKDEL 806

RESULT 9  
 ABG66686  
 ID ABG66686 standard; protein; 806 AA.  
 XX  
 AC ABG66686;  
 XX  
 DT 30-AUG-2002 (first entry)  
 XX  
 DE Human novel polypeptide #21.  
 XX  
 KW Human; inflammatory condition; shock; sepsis; immune response; cancer;  
 KW wound healing; central nervous system disease; haematopoiesis;  
 KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;  
 KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;  
 KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;  
 KW bone degenerative disorder; periodontal disease; reperfusion injury;  
 KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;  
 KW allergic condition; thrombolytic; thrombosis; coagulation disorder;  
 KW fungal infection.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200244340-A2.  
 XX  
 PD 06-JUN-2002.  
 XX  
 PF 30-NOV-2001; 2001WO-US047004.  
 XX  
 PR 30-NOV-2000; 2000US-00728952.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;  
 PI Yamazaki V, Ujwal ML, Drmanac RT;  
 XX  
 DR N-PSDB; ABK94910.  
 XX  
 PT Novel nucleic acids and polypeptides for diagnosis, treatment of  
 PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell  
 PT disorders, cancer and promoting wound healing.  
 XX  
 PS Claim 10; Page 584-585; 672pp; English.  
 XX  
 CC The invention relates to human novel polynucleotides and associated  
 CC polypeptides. The polynucleotides and polypeptides are useful for  
 CC treating inflammatory conditions such as arthritis, nephritis, Crohn's  
 CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses  
 CC and cancer and for promoting wound healing. The sequences are used to  
 CC induce the proliferation of neural cells and regeneration of nerve and  
 CC brain tissue, and are useful for the treatment of central and peripheral  
 CC nervous system diseases and neuropathies, such as Alzheimer's disease,  
 CC Parkinson's disease, Huntington's disease and amyotrophic lateral  
 CC sclerosis. The sequences are involved in chemotactic or chemokinetic  
 CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid  
 CC cell disorders and platelet disorders such as thrombocytopenia,  
 CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue  
 CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of  
 CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal  
 CC disease. The sequences of the invention are also useful for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues, immune deficiencies and disorders  
 CC including severe combined immunodeficiency (SCID), bacterial or fungal  
 CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia  
 CC gravis, allergic conditions such as asthma, thrombolytic or thrombotic  
 CC and coagulation disorders. Sequences ABG66666-ABG66758 represent human  
 CC novel polypeptides of the invention.  
 XX  
 SQ Sequence 806 AA;

Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 580; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 155 KINKLEKAVAAATFFVGNPEHMEMQONLDYYQTMGSKVEADFKDLETPQPHQBEPLGVR 214  
 DB 225 KINKLEKAVAAATFFVGNPEHMEMQONLDYYQTMGSKVEADFKDLETPQPHQBEPLGVR 284  
 QY 215 LYSEBQPOEAVPHLEAALQEVFAVEECRALCEGYDGYNLYEYNADLFOAIDHVIQ 274  
 DB 285 LYSEBQPOEAVPHLEAALQEVFAVEECRALCEGYDGYNLYEYNADLFOAIDHVIQ 344  
 QY 275 VLNCQNCVTELASHPSREKPFEDFLPSHNYLQFAYNIGNYTOAGCAKTYLLFFPND 334  
 DB 345 VLNCQNCVTELASHPSREKPFEDFLPSHNYLQFAYNIGNYTOAGCAKTYLLFFPND 404  
 QY 335 EYMNQNLAYAAMLGEBHTRSGPRESAKYRQSRLLKELLFFAYDVGIFVDPDSWT 394  
 DB 405 EYMNQNLAYAAMLGEBHTRSGPRESAKYRQSRLLKELLFFAYDVGIFVDPDSWT 464  
 QY 395 PEEVIPKRLQEKQKSERETAVRISOEIGNLMKEIETLVEEKTESLDVSRLTREGGPLY 454  
 DB 465 PEEVIPKRLQEKQKSERETAVRISOEIGNLMKEIETLVEEKTESLDVSRLTREGGPLY 524  
 QY 455 EGISLTMSKLLNGYQVRVMDGVIISDHECOELQRLTNVAATSGGYRCQTSPTHTNEKPY 514  
 DB 525 EGISLTMSKLLNGYQVRVMDGVIISDHECOELQRLTNVAATSGGYRCQTSPTHTNEKPY 584  
 QY 515 GYTVFKALKIGOGKVPLQSAHLYNVTKEVRIMESYFRDLTPLYFSYSHLVCRTAEE 574  
 DB 585 GYTVFKALKIGOGKVPLQSAHLYNVTKEVRIMESYFRDLTPLYFSYSHLVCRTAEE 644  
 QY 575 VQAEKDDSHPVHVDNCILNAETLYCVKEPPAYTFRDYSAILYLNGDFDGGNFYFTELDA 634  
 DB 645 VQAEKDDSHPVHVDNCILNAETLYCVKEPPAYTFRDYSAILYLNGDFDGGNFYFTELDA 704  
 QY 635 KTVTAEVQCCGRVGFSSGHTENPHGVKAVTRGQRCALAMFTLDPHRSRDRVQADDLV 694  
 DB 705 KTVTAEVQCCGRVGFSSGHTENPHGVKAVTRGQRCALAMFTLDPHRSRDRVQADDLV 764  
 QY 695 KMLFSPDEMDLSQEQPLDAQQGPPEPAQESLSGSSEKPKDEL 736  
 DB 765 KMLFSPDEMDLSQEQPLDAQQGPPEPAQESLSGSSEKPKDEL 806

RESULT 10  
 AAB36391  
 ID AAB36391 standard; protein; 363 AA.  
 XX  
 AC AAB36391;  
 XX  
 DT 27-FEB-2001 (first entry)  
 XX  
 DE Human tumour suppressor Gros1-L protein SEQ ID NO:2.  
 XX  
 KW Tumour suppressor; Gros1-L; Gros1-S; cell proliferation; regulation;  
 KW cancer; cytostatic; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200065047-A1.  
 XX  
 PD 02-NOV-2000.  
 XX  
 PF 26-APR-2000; 2000WO-JP002731.  
 XX  
 PR 26-APR-1999; 99JP-00118806.  
 XX  
 PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
 XX  
 PI Wadhwa R, Sugihara T, Yoshida A;  
 XX  
 DR WPI; 2000-587340/57.



XX Full-length tumor suppressor genes encoding Grosi-L, Grosi-S and mouse  
PT homologs participating in regulation of cell proliferation, useful in  
PT development of preventives and remedies of cancer.  
XX  
XX  
XX Claim 1; Page 67-69; 114pp; Japanese.  
XX  
XX The present sequence represents the human tumour suppressor designated  
CC Grosi-L. Grosi-L and Grosi-S have cytotostatic activity and can be used in  
CC gene therapy. Grosi-L and Grosi-S genes are useful in the development of  
CC drugs used to treat and prevent cancer  
XX  
SQ Sequence 363 AA;  
  
Query Match 48.9%; Score 360; DB 3; Length 363;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 MAVRALKLLTLLAVVAASQAQAEVESEAGWMTPTDLLFAEGTAAYARGDWPVLSMER 60  
Db 1 MAVRALKLLTLLAVVAASQAQAEVESEAGWMTPTDLLFAEGTAAYARGDWPVLSMER 60  
  
Qy 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
Db 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
  
Qy 121 RRCLGPPAAHSLSEMELEFRKSPYNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180  
Db 121 RRCLGPPAAHSLSEMELEFRKSPYNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180  
  
Qy 181 QNLDDYQTMGSKVEADFKDLETPHMOEFLGVRLYSEEQPOEAVPHLEALQEFYFVAYE 240  
Db 181 QNLDDYQTMGSKVEADFKDLETPHMOEFLGVRLYSEEQPOEAVPHLEALQEFYFVAYE 240  
  
Qy 241 ECRALCEGPDYDGYNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 300  
Db 241 ECRALCEGPDYDGYNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 300  
  
Qy 301 PSHYNYLQFAYNIGNYTOAGECAKTYLLFPNDVNMNQLAYYAAMLGSEHTRSIGPRE 360  
Db 301 PSHYNYLQFAYNIGNYTOAGECAKTYLLFPNDVNMNQLAYYAAMLGSEHTRSIGPRE 360

RESULT 11  
AAB94482  
ID AAB94482 standard; protein; 804 AA.  
XX  
XX AAB94482;  
XX  
XX 26-JUN-2001 (first entry)  
XX  
XX Human protein sequence SEQ ID NO:15160.  
DE  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX  
XX Homo sapiens.  
OS  
XX EP1074617-A2.  
FN  
XX  
XX 07-FEB-2001.  
PD  
XX  
XX 28-JUL-2000; 2000EP-00116126.  
PF  
XX  
XX 29-JUL-1999; 99JP-00248036.  
PR  
XX 27-AUG-1999; 99JP-00300253.  
PR  
XX 11-JAN-2000; 2000JP-00118776.  
PR  
XX 02-MAY-2000; 2000JP-00183767.  
PR  
XX 09-JUN-2000; 2000JP-00241899.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.  
DR  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
PT length cDNAs defined in the specification, and for the detection and/or  
PT diagnosis of the abnormality of the proteins encoded by the full-length  
PT cDNAs.  
XX  
XX Claim 8; SEQ ID NO 15160; 2537pp + Sequence Listing; English.  
PS  
XX  
XX The present invention describes primer sets for synthesizing 5602 full-  
CC length cDNAs defined in the specification. Where a primer set comprises:  
CC (a) an oligo-dr primer and an oligonucleotide complementary to the  
CC complementary strand of a polynucleotide which comprises one of the 5602  
CC nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
CC specification. The primer sets can be used in antisense therapy and in  
CC gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention  
XX  
SQ Sequence 804 AA;  
  
Query Match 38.2%; Score 281; DB 4; Length 804;  
Best Local Similarity 99.4%; Pred. No. 4,7e-262;  
Matches 681; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
Qy 1 MAVRALKLLTLLAVVAASQAQAEVESEAGWMTPTDLLFAEGTAAYARGDWPVLSMER 60  
Db 1 MAVRALKLLTLLAVVAASQAQAEVESEAGWMTPTDLLFAEGTAAYARGDWPVLSMER 60  
  
Qy 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
Db 61 ALRSRAALRALRLRCRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
  
Qy 121 RRCLGPPAAHSLSEMELEFRKSPYNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180  
Db 121 RRCLGPPAAHSLSEMELEFRKSPYNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 180  
  
Qy 181 QNLDDYQTMGSKVEADFKDLETPHMOEFLGVRLYSEEQPOEAVPHLEALQEFYFVAYE 240  
Db 181 QNLDDYQTMGSKVEADFKDLETPHMOEFLGVRLYSEEQPOEAVPHLEALQEFYFVAYE 240  
  
Qy 241 ECRALCEGPDYDGYNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 300  
Db 241 ECRALCEGPDYDGYNYLQVAYFKINKLEKAVAAAHFTFFVGNPEHMEMQ 300  
  
Qy 301 PSHYNYLQFAYNIGNYTOAGECAKTYLLFPNDVNMNQLAYYAAMLGSEHTRSIGPRE 360  
Db 301 PSHYNYLQFAYNIGNYTOAGECAKTYLLFPNDVNMNQLAYYAAMLGSEHTRSIGPRE 360  
  
Qy 361 SAKERYQRSLLEKELLFFAYDVGIPFVDPDPSWTPEEVIKPELOEKQKSERETAVRISGE 420  
Db 361 SAKERYQRSLLEKELLFFAYDVGIPFVDPDPSWTPEEVIKPELOEKQKSERETAVRISGE 420  
  
Qy 421 IGNLMKEIETLVEEKTESLDVSRLTREGGPLLLEGISLTMSKLANGYQRYVMDGVISD 480  
Db 421 IGNLMKEIETLVEEKTESLDVSRLTREGGPLLLEGISLTMSKLANGYQRYVMDGVISD 480  
  
Qy 481 HECQELQRLTNVAASGDCYRGQTSPTHTNEFYGVTVFKALKIQEGKVPLQSAHLYN 540  
Db 481 HECQELQRLTNVAASGDCYRGQTSPTHTNEFYGVTVFKALKIQEGKVPLQSAHLYN 540

Db 481 HECQELQRLTNVAATSGDYGRCQTSPTNPKFYGVTVFKALKLGQEGKVPLOSAHLYN 540  
QY 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPVHVDNCILNAETLVC 600  
Db 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPVHVDNCILNAETLVC 600  
QY 601 VKEPPAYTRDYSAIYLNGDFDGNFYFTELDKTVTAEVQPOCGRAVGFSSGTENPHG 660  
Db 601 VKEPPAYTRDYSAIYLNGDFDGNFYFTELDKTVTAEVQPOCGRAVGFSSGTENPHG 660  
QY 661 VKAVTRGQRCALWFTLDPHRSR 685  
Db 661 VKAVTRGQRCALWFTLDPHRSR 685

RESULT 12  
AAU09860  
ID AAU09860 standard; protein; 708 AA.  
XX  
AC AAU09860;  
XX  
DT 26-FEB-2002 (first entry)  
XX  
DE Novel human secreted protein #1.  
XX  
KW Secreted protein; cytostatic; immunosuppressive; vulnery; vaccine;  
KW antiinflammatory; neuroprotective; nephrotropic; cardiovascular; human;  
KW cancer; autoimmune disease; wound healing disorder; infectious;  
KW haematopoietic disorder; inflammatory disorder; infertility;  
KW neurological disease; psychiatric disease; cardiovascular disease;  
KW respiratory disease; renal; gastrointestinal.  
XX  
OS Homo sapiens.  
XX  
PN WO200179454-A1.  
XX  
PD 25-OCT-2001.  
XX  
PF 11-APR-2001; 2001WO-US011797.  
XX  
PR 13-APR-2000; 2000US-0196603P.  
PR 24-APR-2000; 2000US-0199417P.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
XX (SMIK ) SMITHKLINE BEECHAM PLC.  
XX  
PI Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z;  
XX  
DR WPI; 2002-061975/08.  
DR N-FSDB; AAS17572.  
XX  
PT New secreted proteins or polypeptides, useful for treating e.g. cancer,  
PT autoimmune diseases, wound healing disorder, infections, haematopoietic  
PT disorders, inflammatory disorders, infertility, neurological and  
XX psychiatric diseases, cardiovascular diseases, respiratory diseases,  
XX renal diseases, or gastrointestinal diseases. These may also be used to  
XX treat diseases, abnormalities and disorders caused by abnormal  
XX expression, production, function and/or metabolism of the genes, as  
XX vaccines for inducing immunological response in a mammal, and in  
XX screening methods for detecting the effect of added compounds on the  
XX production of mRNA and polypeptide in cells. The polypeptides can be used  
XX as immunogens to produce antibodies immunospecific for the polypeptides,  
XX and to identify membrane-bound or soluble receptors. The polynucleotides  
XX may be used as diagnostic reagents, in chromosome localisation studies,  
XX and in tissue expression studies. The present sequence represents the  
XX amino acid sequence of novel human secreted protein #1

XX SQ Sequence 708 AA;  
Query Match 37.4%; Score 275; DB 5; Length 708;  
Best Local Similarity 99.7%; Pred. No. 2.6e-256;  
Matches 375; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 361 SAKYRORSLLKEILLFRAYDVFGIPFVDPDSWTPEEIPKLOKQKSERETAVRISQE 420  
Db 333 SAKYRQRSLLKEILLFRAYDVFGIPFVDPDSWTPEEIPKLOKQKSERETAVRISQE 392  
QY 421 IGNLMKEIETLVEEKTKESLDVSRLTREGGPLLVEGISLTNNSKLLNGYQVRVMDGVISD 480  
Db 393 IGNLMKEIETLVEEKTKESLDVSRLTREGGPLLVEGISLTNNSKLLNGYQVRVMDGVISD 452  
QY 481 HECQELQRLTNVAATSGDYGRCQTSPTNPKFYGVTVFKALKLGQEGKVPLOSAHLYN 540  
Db 453 HECQELQRLTNVAATSGDYGRCQTSPTNPKFYGVTVFKALKLGQEGKVPLOSAHLYN 512  
QY 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPVHVDNCILNAETLVC 600  
Db 513 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEEVOAERKDDSHPVHVDNCILNAETLVC 572  
QY 601 VKEPPAYTRDYSAIYLNGDFDGNFYFTELDKTVTAEVQPOCGRAVGFSSGTENPHG 660  
Db 573 VKEPPAYTRDYSAIYLNGDFDGNFYFTELDKTVTAEVQPOCGRAVGFSSGTENPHG 632  
QY 661 VKAVTRGQRCALWFTLDPHRSRDRVQADDLVKMLFSPEEMDLSQBPDLAQOQGPPEP 720  
Db 633 VKAVTRGQRCALWFTLDPHRSRDRVQADDLVKMLFSPEEMDLSQBPDLAQOQGPPEP 692  
QY 721 AQESLSGSESKEPKDEL 736  
Db 693 AQESLSGSESKEPKDEL 708

RESULT 13  
AAE04246  
ID AAE04246 standard; protein; 337 AA.  
XX  
AC AAE04246;  
XX  
DT 09-AUG-2001 (first entry)  
XX  
DE Human gene 2 encoded secreted protein fragment, SEQ ID NO:103.  
XX  
KW Human; secreted protein; proliferative disorder; cancer; tumour;  
KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KW skin disorder; psoriasis; sepsis; diabetes; kidney disorder;  
KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;  
KW endocrine disorder; infection; wound healing; vulnery; cell culture;  
KW chemotaxis; food additive; binding partner identification.  
XX  
OS Homo sapiens.  
XX  
PN WO200136432-A2.  
XX  
PD 25-MAY-2001.  
XX  
PF 15-NOV-2000; 2000WO-US031162.  
XX  
PR 19-NOV-1999; 99US-0166415P.  
PR 30-JUN-2000; 2000US-0215136P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Komatsoulis GA, Baker KP, Young PE;  
XX WPI; 2001-343793/36.  
DR

XX Isolated nucleic acid molecule encoding a human secreted protein is used  
PT in preventing, treating or ameliorating a medical condition.  
XX  
XX Disclosure; Page 12; 509pp; English.  
XX  
XX AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted  
CC protein genes, and AAE04199-AAE04239 represent the proteins they encode.  
CC AAD04240-AAE04297 represent human secreted protein fragments or variants.  
CC The secreted proteins and their genes are useful for preventing, treating  
CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
CC Pathological conditions can be diagnosed by determining the amount of the  
CC new protein in a sample or by determining the presence of mutations in  
CC the new genes. Specific uses are described for each of the 18 genes,  
CC based on the tissues in which they are most highly expressed, and include  
CC developing products for the diagnosis or treatment of proliferative  
CC disorders, cancer, tumors, foetal and developmental abnormalities,  
CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
CC angiogenic disorders, kidney disorders, gastrointestinal disorders, the  
CC pregnancy-related disorders, endocrine disorders, and infections. The  
CC proteins can also be used to aid wound healing and epithelial cell  
CC proliferation, to prevent skin ageing due to sunburn, to maintain organs  
CC before transplantation, for supporting cell culture of primary tissues,  
CC to regenerate tissues, to identify their cognate ligands or binding  
CC partners, and in chemoradiotherapy, and can be used as a food additive or  
CC preservative to modify storage properties. Antibodies specific for a  
CC protein of the invention can be used in alleviating symptoms associated  
CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
CC present sequence represents a human secreted protein fragment referred to  
CC in the disclosure of the invention  
XX  
XX SQ Sequence 337 AA;

Query Match 36.3%; Score 267; DB 4; Length 337;  
Best Local Similarity 100.0%; Pred. No. 7.3e-249;  
Matches 267; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 470 QRVMDGVISDHECOELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLGQEGK 529  
DB 71 QRVMDGVISDHECOELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLGQEGK 130  
QY 530 VPLQSAHLYNVNTEKVRIMESYRLDTPLYFSYSHLVCRTAIEVQAEKDDSHPVHVD 589  
DB 131 VPLQSAHLYNVNTEKVRIMESYRLDTPLYFSYSHLVCRTAIEVQAEKDDSHPVHVD 190  
QY 590 NCILNAETLVCKBPPAYTFRDYSAILYLNGDFDGGNFYFTLDAKTVAEVQPCGRV 649  
DB 191 NCILNAETLVCKBPPAYTFRDYSAILYLNGDFDGGNFYFTLDAKTVAEVQPCGRV 250  
QY 650 GFSSGTENPHGVKAVTRGQRCIAIWLFTLDRHSDRDRVQADDLVKMLFSPPEMDLSQEQ 709  
DB 251 GFSSGTENPHGVKAVTRGQRCIAIWLFTLDRHSDRDRVQADDLVKMLFSPPEMDLSQEQ 310  
QY 710 PLDAQQGPPEPAQESLSGSESKPKDEL 736  
DB 311 PLDAQQGPPEPAQESLSGSESKPKDEL 337

RESULT 14  
ABP69128  
ID ABP69128 standard; protein; 706 AA.  
XX  
XX AC ABP69128;  
XX  
DT 20-JAN-2003 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 1175.  
XX

KW Human: genome mapping; gene therapy; food supplement; virus; fungus;  
KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
KW arthritis; cytostatic; immunomodulator; neurotropic; dermatological;  
KW antiparkinsonian; antidiabetic; immunosuppressive; haematological;  
KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
KW antiarthritic.  
XX Homo sapiens.  
OS  
XX  
XX WO200270539-A2.  
PN  
XX 12-SEP-2002.  
PD  
XX 05-MAR-2002; 2002WO-US005095.  
PF  
XX 05-MAR-2001; 2001US-00799451.  
PR  
XX (HYSE-) HYSEQ INC.  
PA  
XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
PI Wehrman T, Wang J, Wang D, Drmanac RT;  
XX WPI; 2002-759812/82.  
DR N-PSDB; AB211345.  
XX  
XX New polynucleotides comprising sequences assembled from expressed  
PT sequence tags (ESTs), useful for treating cell-proliferative, or platelet  
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet  
PT or coagulation disorders.  
XX  
XX Claim 9; SEQ ID NO 1175; 1012pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated polynucleotide (T) comprising a  
CC nucleotide sequence selected from any of 948 sequences (AB211119-  
CC AB212066) or their mature protein coding portion, active domain coding  
CC protein or complementary sequences. The polynucleotides are useful for  
CC identifying expressed genes or for physical mapping of human genome. The  
CC encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight  
CC markers, as a food supplement, for generating antibodies, in medical  
CC imaging, screening and diagnostic assays and for treating cell-  
CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's  
CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,  
CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,  
CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver  
CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),  
CC arthritis, etc. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX SQ Sequence 706 AA;

Query Match 33.2%; Score 244; DB 5; Length 706;  
Best Local Similarity 100.0%; Pred. No. 2.3e-226;  
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 493 AATSGDGYRGQTSPTNEKFGYGVTVFKALKLGQEGKVPLOSAHLYNVNTEKVRIMESY 552  
DB 463 AATSGDGYRGQTSPTNEKFGYGVTVFKALKLGQEGKVPLOSAHLYNVNTEKVRIMESY 522  
QY 553 FRLDTPLYFSYSHLVCRTAIEVQAEKDDSHPVHVDNCILNAETLVCKBPPAYTFRDY 612  
DB 523 FRLDTPLYFSYSHLVCRTAIEVQAEKDDSHPVHVDNCILNAETLVCKBPPAYTFRDY 582  
QY 613 SAILYLNGDFDGGNFYFTLDAKTVAEVQPCGRVAVGSSGTENPHGVKAVTRGQRCI 672  
DB 583 SAILYLNGDFDGGNFYFTLDAKTVAEVQPCGRVAVGSSGTENPHGVKAVTRGQRCI 642  
QY 673 ALWFTLDRHSDRDRVQADDLVKMLFSPPEMDLSQEQPLDAQQGPPEPAQESLSGSESKP 732  
DB 643 ALWFTLDRHSDRDRVQADDLVKMLFSPPEMDLSQEQPLDAQQGPPEPAQESLSGSESKP 702

QY 733 KDEL 736  
DB 703 KDEL 706

RESULT 15  
AAE04200  
ID AAE04200 standard; protein; 359 AA.  
AC AAE04200;  
XX  
XX 09-AUG-2001 (first entry)  
XX  
XX Human gene 2 encoded secreted protein HUVFY29, SEQ ID NO:54.  
DE  
XX Human; secreted protein; proliferative disorder; cancer; tumour;  
KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;  
KW endocrine disorder; infection; wound healing; vulnerability; cell culture;  
KW chemotaxis; food additive; binding partner identification.  
XX  
OS Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FT Peptide 1..18  
FT Protein 19..359  
FT /label= signal\_peptide  
FT /note= "Mature secreted protein"

XX WO200136432-A2.  
XX  
PD 25-MAY-2001.  
XX  
XX 15-NOV-2000; 2000WO-US031162.  
XX  
XX 19-NOV-1999; 99US-0166415P.  
XX 30-JUN-2000; 2000US-0215136P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Ruben SM, Komatsoulis GA, Baker KP, Young PE;  
XX WPI; 2001-343793/36.  
XX N-PSDB; AAD08489.  
XX  
XX Isolated nucleic acid molecule encoding a human secreted protein is used  
XX in preventing, treating or ameliorating a medical condition.  
XX  
XX Claim 11; Page 431-433; 509pp; English.  
XX  
XX AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted  
XX protein genes, and AAE04199-AAE04239 represent the proteins they encode.  
XX AAE04240-AAE04297 represent human secreted protein fragments or variants.  
XX The secreted proteins and their genes are useful for preventing, treating  
XX or ameliorating medical conditions, e.g., by protein or gene therapy.  
XX Pathological conditions can be diagnosed by determining the amount of the  
XX new protein in a sample or by determining the presence of mutations in  
XX the new genes. Specific uses are described for each of the 18 genes,  
XX based on the tissues in which they are most highly expressed, and include  
XX developing products for the diagnosis or treatment of proliferative  
XX disorders, cancer, tumours, foetal and developmental abnormalities,  
XX haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
XX diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
XX neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
XX cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
XX psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
XX angiogenic disorders, kidney disorders, gastrointestinal disorders,

CC pregnancy-related disorders, endocrine disorders, and infections. The  
CC proteins can also be used to aid wound healing and epithelial cell  
CC proliferation, to prevent skin ageing due to sunburn, to maintain organs  
CC before transplantation, for supporting cell culture of primary tissues,  
CC to regenerate tissues, to identify their cognate ligands or binding  
CC partners, and in chemotaxis, and can be used as a food additive or  
CC preservative to modify storage properties. Antibodies specific for a  
CC protein of the invention can be used in alleviating symptoms associated  
CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
XX present sequence represents a human secreted protein of the invention  
XX  
SQ Sequence 359 AA;

Query Match 29.8%; Score 219; DB 4; Length 359;  
Best Local Similarity 99.7%; Pred. No. 1.8e-202;  
Matches 319; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MAVRALKLTLLAVVAASQAEVESEAGWMTVDLLFAEGTAAYARGDWGVLSMER 60  
DB 1 MAVRALKLTLLAVVAASQAEVESEAGWMTVDLLFAEGTAAYARGDWGVLSMER 60  
QY 61 ALRSRAALRALRLRCRTQCAADFPWELDPDPSPAQASGAGALRDLSPFGGLIRRAACL 120  
DB 61 ALRSRAALRALRLRCRTQCAADFPWELDPDPSPAQASGAGALRDLSPFGGLIRRAACL 120  
QY 121 RRCLGPPAAHSLSEEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHFFVGNPEHMEMQ 180  
DB 121 RRCLGPPAAHSLSEEMELEFRKSPYNYLOVAYFKINKLEKAVAAAHFFVGNPEHMEMQ 180  
QY 181 QNLDYYQTMGSKVEADFXDLETPQPMQBRFLGRVRLYSEEQPOEAVPHLEALQYFVAYE 240  
DB 181 QNLDYYQTMGSKVEADFXDLETPQPMQBRFLGRVRLYSEEQPOEAVPHLEALQYFVAYE 240  
QY 241 ECRALCEGPDYDGYNYLEYNADLFQAITDHYIQLVNLCKQNCVTELASHSPREKPFDFL 300  
DB 241 ECRALCEGPDYDGYNYLEYNADLFQAITDHYIQLVNLCKQNCVTELASHSPREKPFDFL 300  
QY 301 PSHYNYLQFAYYNYGNYTQA 320  
DB 301 PSHYNYLQFAYYNYGNYTQA 320

Search completed: July 18, 2004, 09:58:11  
Job time : 64 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 09:30:33 ; Search time 87 Seconds  
(without alignments)  
2669.211 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 3870  
Sequence: 1 MAVRALKLTLLAVVAAS.....PPEPAQESLSGSESKPKDEL 736

Scoring table: BLOSUM62  
Gap 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25: \*  
1: sp\_archaea: \*  
2: sp\_bacteria: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mnc: \*  
8: sp\_organelle: \*  
9: sp\_phase: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp\_vertebrate: \*  
14: sp\_unclassified: \*  
15: sp\_rvrius: \*  
16: sp\_bacteriap: \*  
17: sp\_archaeap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3870	100.0	736	4 Q9HC86	Q9HC86 homo sapien
2	3846	99.4	736	4 Q96SK8	Q96SK8 homo sapien
3	3844	99.3	736	4 Q96SL5	Q96SL5 homo sapien
4	3582	92.6	804	4 Q96SN3	Q96SN3 homo sapien
5	3403.5	87.9	739	11 Q8CFL3	Q8CFL3 mus musculus
6	3379	87.3	738	11 Q8BSV2	Q8BSV2 mus musculus
7	3378.5	87.3	728	11 Q9RLJ8	Q9RLJ8 rattus norv
8	3316	85.7	747	11 Q9QZT7	Q9QZT7 mus musculus
9	3305	85.4	742	11 Q9CWX5	Q9CWX5 mus musculus
10	2384	61.6	542	11 Q9QZT6	Q9QZT6 mus sapien
11	2024	52.3	390	4 Q96BR8	Q96BR8 homo sapien
12	2014	52.0	390	4 Q9H6K3	Q9H6K3 homo sapien
13	1911	49.4	363	4 Q9HC87	Q9HC87 homo sapien
14	1747	45.1	708	4 Q8IVL5	Q8IVL5 homo sapien
15	1715.5	44.3	703	11 Q8CQ71	Q8CQ71 mus musculus
16	1473	38.1	736	4 Q8IVL6	Q8IVL6 homo sapien

17	1421.5	36.7	732	11 Q8CG70	Q8CG70 mus musculus
18	1409.5	36.4	527	4 Q9NV12	Q9NV12 homo sapien
19	1179.5	30.5	551	4 Q13512	Q13512 homo sapien
20	1179.5	30.5	551	4 Q15740	Q15740 homo sapien
21	1127.5	29.1	545	11 Q88836	Q88836 mus musculus
22	1106.5	28.6	409	11 Q8C673	Q8C673 mus musculus
23	672	17.4	412	5 Q9NDP6	Q9NDP6 ciona intes
24	554.5	14.3	388	4 Q8NB01	Q8NB01 homo sapien
25	536.5	13.9	400	11 Q8C8C5	Q8C8C5 mus musculus
26	498	12.9	443	11 Q8K2B0	Q8K2B0 mus musculus
27	474.5	12.3	457	11 Q8C112	Q8C112 mus musculus
28	336	8.7	476	5 Q9NA33	Q9NA33 caenorhabdi
29	173.5	4.5	315	11 Q9D1E6	Q9D1E6 mus musculus
30	163	4.2	319	4 Q9H6J2	Q9H6J2 homo sapien
31	144	3.7	383	10 Q9C9X2	Q9C9X2 arabidopsis
32	144	3.7	1025	2 Q84IL2	Q84IL2 clostridium
33	141	3.6	342	16 Q67735	Q67735 aquifex aeo
34	135	3.5	1270	4 Q96JN2	Q96JN2 homo sapien
35	132	3.4	197	16 Q8DKV0	Q8DKV0 synchococc
36	129	3.3	227	4 Q13513	Q13513 homo sapien
37	127.5	3.3	1711	17 Q8Q0F8	Q8Q0F8 methanosarc
38	126.5	3.2	1075	5 Q8WQ25	Q8WQ25 leishmania
39	125.5	3.2	1033	4 Q9UF81	Q9UF81 homo sapien
40	125.5	3.2	1440	3 Q871S3	Q871S3 neurospora
41	123.5	3.2	640	17 Q9V0S2	Q9V0S2 pyrococcus
42	123	3.2	1309	16 Q8CX10	Q8CX10 streptococc
43	123	3.2	1323	16 Q8E5G4	Q8E5G4 streptococc
44	122.5	3.2	1831	10 Q7XN11	Q7XN11 oryza sativ
45	122	3.2	1529	10 Q81068	Q81068 arabidopsis

ALIGNMENTS

RESULT 1

ID	Q9HC86	PRELIMINARY;	PRT;	736 AA.
AC	Q9HC86;			
DT	01-MAR-2001 (TREMELrel. 16, Created)			
DT	01-MAR-2001 (TREMELrel. 16, Last sequence update)			
DE	01-OCT-2003 (TREMELrel. 25, Last annotation update)			
DE	GROSL-L protein.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-Testis;			
RX	MEDLINE=20406537; PubMed=10951563;			
RA	Kaul S.C., Sugihara T., Yoshida A., Nomura H., Wadhwa R.;			
RT	Grosi, a potential growth suppressor on chromosome 1: its identity to			
RL	baseament membrane-associated, proteoglycan, leprecan."			
DR	EMBL; AF097432; AAC31019.1;			
DR	GO; GO:0016706; P:oxidoreductase activity, acting on paired d. . ; IEA.			
DR	GO; GO:001938; P:protein metabolism; IEA.			
DR	InterPro; IPR005123; 2OG-Fell_Oxy.			
DR	InterPro; IPR000886; ER_target_S.			
DR	InterPro; IPR006620; PR_4_hyd_alph.			
DR	InterPro; IPR008941; TPR-like.			
DR	Pfam; PF03171; 2OG-Fell_Oxy; 1.			
DR	SMART; SM00702; P4HG; 1.			
DR	PROSITE; PS00014; ER_TARGET; 1.			
SQ	SEQUENCE 736 AA; 83413 MW; 59D1205DB76ADB50 CRC64;			

Query Match 100.0%; Score 3870; DB 4; Length 736;  
Best Local Similarity 100.0%; Pred. No. 1.3e-282;  
Matches 736; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAVRALKLTLLAVVAASQAEVSEAGWGVTPTDLLFAEGTAAYARGDVGVLWNER 60  
DB 1 MAVRALKLTLLAVVAASQAEVSEAGWGVTPTDLLFAEGTAAYARGDVGVLWNER 60

QY 61 ALRSSAALRALRLRCLRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
 DB 61 ALRSSAALRALRLRCLRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
 QY 121 RRLGPPAAHLSISEMELEFRKSPYNYLQVAYFKINKLEKAVAAAHTEFFVGNPEHMEMQ 180  
 DB 121 RRLGPPAAHLSISEMELEFRKSPYNYLQVAYFKINKLEKAVAAAHTEFFVGNPEHMEMQ 180  
 QY 181 QNLDDYQTMGSGVKEADFKDLTQPHMQEFLGVRLYSEEQPQAVPHLEALQEFYFVAYE 240  
 DB 181 QNLDDYQTMGSGVKEADFKDLTQPHMQEFLGVRLYSEEQPQAVPHLEALQEFYFVAYE 240  
 QY 241 ECRALCEGPDYDGVNLYEYNADLFOAITDHYIQLVCKQNCVTELASHSPREKPFDFL 300  
 DB 241 ECRALCEGPDYDGVNLYEYNADLFOAITDHYIQLVCKQNCVTELASHSPREKPFDFL 300  
 QY 301 PSHYNYLOFAYNIGNYTQAGECAKTYLLFPFNDVNNQNLAYYAAMLGSEHTRSIGPRE 360  
 DB 301 PSHYNYLOFAYNIGNYTQAGECAKTYLLFPFNDVNNQNLAYYAAMLGSEHTRSIGPRE 360  
 QY 361 SAKYRQBSLLEKELLFPAYDVFGIPFVDPDSWTPPEVPIKRLQEKOKSERETAVRISQE 420  
 DB 361 SAKYRQBSLLEKELLFPAYDVFGIPFVDPDSWTPPEVPIKRLQEKOKSERETAVRISQE 420  
 QY 421 IGLMKETETLVEEKTRESLDVSRITREGGPLLVEGISTWNSKLLNGYQVWMDGVISD 480  
 DB 421 IGLMKETETLVEEKTRESLDVSRITREGGPLLVEGISTWNSKLLNGYQVWMDGVISD 480  
 QY 481 HECQELQRLTNVAATSGDGYRGQTSPTNEKFGYVTVFKALKGQSKVPLOSALHYNN 540  
 DB 481 HECQELQRLTNVAATSGDGYRGQTSPTNEKFGYVTVFKALKGQSKVPLOSALHYNN 540  
 QY 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEBQVQERKDDSHPVHVDNCILNAETLVC 600  
 DB 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEBQVQERKDDSHPVHVDNCILNAETLVC 600  
 QY 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTELDQKTYTAEVQPCGRAVGFSSGTENPHG 660  
 DB 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTELDQKTYTAEVQPCGRAVGFSSGTENPHG 660  
 QY 661 VKAVTRGQRCALALWFTLDRHSRDRVQADLLVKMLFSPPEMDLSQEQPLDAQCGGPPPEP 720  
 DB 661 VKAVTRGQRCALALWFTLDRHSRDRVQADLLVKMLFSPPEMDLSQEQPLDAQCGGPPPEP 720  
 QY 721 AQESLSGSESXPDEL 736  
 DB 721 AQESLSGSESXPDEL 736

RESULT 2  
 Q96SK8 ID Q96SK8 PRELIMINARY; PRT; 736 AA.  
 AC Q96SK8  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein FLJ14791 (Hypothetical protein NT2RP2003671).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,  
 RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,  
 RA Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;  
 RT "NEDO human cDNA sequencing project";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.

RA Ota T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayashi K., Ishii S.,  
 RA Saito K., Yamamoto J., Wakamatsu A., Nagai T., Nakamura Y.,  
 RA Nagahari K., Sugano S., Isogai T.;  
 RT "HRI human cDNA sequencing project";  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK027697; BAB55305.1; -;  
 DR EMBL; AK075418; BAC11608.1; -;  
 DR InterPro; IPR005123; 2OG-Fell\_Oxy.  
 DR InterPro; IPR000886; ER-target\_S.  
 DR InterPro; IPR008941; TPR-like.  
 DR Pfam; PF03171; 2OG-Fell\_Oxy; 1.  
 DR PROSITE; PS00014; ER-TARGET; 1.  
 KW Hypothetical protein\_  
 SQ SEQUENCE 736 AA; 83393 MW; EA1909828FAB685E CRC64;  
 Query Match 99.4%; Score 3846; DB 4; Length 736;  
 Best Local Similarity 99.6%; Pred. No. 8.6e-281;  
 Matches 733; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 MAVRALKLLTLLAVVAASQAEVSEAGWGTPTDLLFAEGTAAYARGDMPGVLSMER 60  
 DB 1 MAVRALKLLTLLAVVAASQAEVSEAGWGTPTDLLFAEGTAAYARGDMPGVLSMER 60  
 QY 61 ALRSSAALRALRLRCLRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
 DB 61 ALRSSAALRALRLRCLRTQCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120  
 QY 121 RRLGPPAAHLSISEMELEFRKSPYNYLQVAYFKINKLEKAVAAAHTEFFVGNPEHMEMQ 180  
 DB 121 RRLGPPAAHLSISEMELEFRKSPYNYLQVAYFKINKLEKAVAAAHTEFFVGNPEHMEMQ 180  
 QY 181 QNLDDYQTMGSGVKEADFKDLTQPHMQEFLGVRLYSEEQPQAVPHLEALQEFYFVAYE 240  
 DB 181 QNLDDYQTMGSGVKEADFKDLTQPHMQEFLGVRLYSEEQPQAVPHLEALQEFYFVAYE 240  
 QY 241 ECRALCEGPDYDGVNLYEYNADLFOAITDHYIQLVCKQNCVTELASHSPREKPFDFL 300  
 DB 241 ECRALCEGPDYDGVNLYEYNADLFOAITDHYIQLVCKQNCVTELASHSPREKPFDFL 300  
 QY 301 PSHYNYLOFAYNIGNYTQAGECAKTYLLFPFNDVNNQNLAYYAAMLGSEHTRSIGPRE 360  
 DB 301 PSHYNYLOFAYNIGNYTQAGECAKTYLLFPFNDVNNQNLAYYAAMLGSEHTRSIGPRE 360  
 QY 361 SAKYRQBSLLEKELLFPAYDVFGIPFVDPDSWTPPEVPIKRLQEKOKSERETAVRISQE 420  
 DB 361 SAKYRQBSLLEKELLFPAYDVFGIPFVDPDSWTPPEVPIKRLQEKOKSERETAVRISQE 420  
 QY 421 IGLMKETETLVEEKTRESLDVSRITREGGPLLVEGISTWNSKLLNGYQVWMDGVISD 480  
 DB 421 IGLMKETETLVEEKTRESLDVSRITREGGPLLVEGISTWNSKLLNGYQVWMDGVISD 480  
 QY 481 HECQELQRLTNVAATSGDGYRGQTSPTNEKFGYVTVFKALKGQSKVPLOSALHYNN 540  
 DB 481 HECQELQRLTNVAATSGDGYRGQTSPTNEKFGYVTVFKALKGQSKVPLOSALHYNN 540  
 QY 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEBQVQERKDDSHPVHVDNCILNAETLVC 600  
 DB 541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEBQVQERKDDSHPVHVDNCILNAETLVC 600  
 QY 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTELDQKTYTAEVQPCGRAVGFSSGTENPHG 660  
 DB 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTELDQKTYTAEVQPCGRAVGFSSGTENPHG 660  
 QY 661 VKAVTRGQRCALALWFTLDRHSRDRVQADLLVKMLFSPPEMDLSQEQPLDAQCGGPPPEP 720  
 DB 661 VKAVTRGQRCALALWFTLDRHSRDRVQADLLVKMLFSPPEMDLSQEQPLDAQCGGPPPEP 720  
 QY 721 AQESLSGSESXPDEL 736  
 DB 721 AQESLSGSESXPDEL 736

Q96SL5	PRELIMINARY;	PRT;	736 AA.	
ID	Q96SL5			
AC	Q96SL5			
DT	01-DEC-2001 (TREMELrel. 19, Created)			
DT	01-DEC-2001 (TREMELrel. 19, Last sequence update)			
DT	01-OCT-2003 (TREMELrel. 25, Last annotation update)			
DE	Hypothetical protein FLJ14774.			
OS	Homo sapiens (human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,			
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,			
RA	Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,			
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,			
RA	Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,			
RA	Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,			
RA	Ninomiya K., Iwayanagi T.,			
RT	"NEO human cDNA sequencing project";			
RL	Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AK027648; BAB55264.1; -			
DR	Genew; HGNC:19316; LEPREL.			
DR	GO; GO:0016706; P:oxidoreductase activity, acting on paired d. . .; IEA.			
DR	GO; GO:0019538; P:protein metabolism; IEA.			
DR	InterPro; IPR005123; 2OG-Fell_Oxy.			
DR	InterPro; IPR006620; Pro_4_hyd_alph.			
DR	InterPro; IPR008941; TPR-like.			
DR	Pfam; PF03171; 2OG-Fell_Oxy; 1.			
DR	SMART; SM00702; P4HC; 1.			
DR	SMART; SM00702; P4HC; 1.			
KW	Hypothetical protein.			
SQ	SEQUENCE 736 AA; 83411 MW; EAL819938FAE695E CRC64;			
Query Match	99.3%; Score 3844; DB 4; Length 736;			
Best Local Similarity	99.3%; Pred. No. 1.2e-280;			
Matches	733; Conservative 1; Mismatches 3; Indels 0; Gaps 0;			
QY	1 MAVRALKLLTLLAVVAASQAQAEVESEAGWGVTDDLFAEGTAAAYARGDWPVGLSMER 60			
DB	1 MAVRALKLLTLLAVVAASQAQAEVESEAGWGVTDDLFAEGTAAAYARGDWPVGLSMER 60			
QY	61 ALRSRAALRALRLCRTOCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120			
DB	61 ALRSRAALRALRLCRTOCAADFPWELDPDWSPPAQASGAGALRDLSPFGGLLRAACL 120			
QY	121 RCLGPPAAHLSSEMELEFRKRSYNYLQVAYFKINKLEKAVAAAHFFVGNPEHMEQ 180			
DB	121 RCLGPPAAHLSSEMELEFRKRSYNYLQVAYFKINKLEKAVAAAHFFVGNPEHMEQ 180			
QY	181 QNLDYQTMGSVKAEADFKDLETPHMQEFLRGVRLYSEEQPOEAVPHLEALQEFVAYE 240			
DB	181 QNLDYQTMGSVKAEADFKDLETPHMQEFLRGVRLYSEEQPOEAVPHLEALQEFVAYE 240			
QY	241 ECRALCEGPDYDGYNYLQVAYFKINKLEKAVAAAHFFVGNPEHMEQ 300			
DB	241 ECRALCEGPDYDGYNYLQVAYFKINKLEKAVAAAHFFVGNPEHMEQ 300			
QY	301 PSHYNYLOFAYNIGNYTOAGECAKTYLLFPDNDVNNQNLAYYAMLGEHTRSGPRE 360			
DB	301 PSHYNYLOFAYNIGNYTOAGECAKTYLLFPDNDVNNQNLAYYAMLGEHTRSGPRE 360			
QY	361 SAKERYORSLEKELLFFAYDVFQIPVDPSDTPTEEVIPKRLQEKQKSERETAVRISQ 420			
DB	361 SAKERYORSLEKELLFFAYDVFQIPVDPSDTPTEEVIPKRLQEKQKSERETAVRISQ 420			
QY	421 IGMLKBIETLVEKTKESLDVSLRTREGGPLLYEGISLTWNSKLLNGYORVMDGVISD 480			
DB	421 IGMLKBIETLVEKTKESLDVSLRTREGGPLLYEGISLTWNSKLLNGYORVMDGVISD 480			
QY	481 HECCELQRLTNVAATSGDVGRTQSPHTPNEKFGVTVFKALKLQBGKVPLOSALHYN 540			
DB	481 HECCELQRLTNVAATSGDVGRTQSPHTPNEKFGVTVFKALKLQBGKVPLOSALHYN 540			
QY	541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAERKDDSHPVHVDNCILNAETLVC 600			
DB	541 VTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEVQAERKDDSHPVHVDNCILNAETLVC 600			

```

Db 241 ECRCALCEGYDYGNYLYNADLFOAITDHYIQLVNCNKQNCVTELASHPSREKPFEDFL 300
Qy 301 PSHVNYLQFAYNIGNYTOAGCACTYLLFFPNDEVMNQNLAYYAAMLGEEHTRISIGPRE 360
Db 301 PSHVNYLQFAYNIGNYTOAGCACTYLLFFPNDEVMNQNLAYYAAMLGEEHTRISIGPRE 360
Qy 361 SAKVQRORSLEKEALLFAYDVGIGPFVDPDSWTPPEVPIKGLQEKOKSERETAVRISQE 420
Db 361 SAKVQRORSLEKEALLFAYDVGIGPFVDPDSWTPPEVPIKGLQEKOKSERETAVRISQE 420
Qy 421 IGNLMKEIETLVEEKTESLVSRLTREGGPLYEGISLTMNSKLLNGYQRVVMGDVISD 480
Db 421 IGNLMKEIETLVEEKTESLVSRLTREGGPLYEGISLTMNSKLLNGYQRVVMGDVISD 480
Qy 481 HECQELQRLTNVAATSGDGYRGQTSPTPNKFKYGVTVFKALKLQEGKVPLOSAHLIYN 540
Db 481 HECQELQRLTNVAATSGDGYRGQTSPTPNKFKYGVTVFKALKLQEGKVPLOSAHLIYN 540
Qy 541 VTEKVRIRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKODSDHPVHVDNCILNAETLVC 600
Db 541 VTEKVRIRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKODSDHPVHVDNCILNAETLVC 600
Qy 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAEVQPCGRAVGFSSGTENPHG 660
Db 601 VKEPPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAEVQPCGRAVGFSSGTENPHG 660
Qy 661 VKAVTRGORCAIALWFTLDRHRSERDR 687
Db 661 VKAVTRGORCAIALWFTLDRHRSERDR 687

RESULT 5
Q8CFL3 PRELIMINARY; PRT; 739 AA.
AC Q8CFL3
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Similar to leprecan.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC024047; AAH24047.1; -.
DR PIR; PT0551.
DR GO; GO:0016706; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0019538; P:protein metabolism; IEA.
DR InterPro; IPR005123; 2OG-FeII_Oxy.
DR InterPro; IPR000886; ER target S.
DR InterPro; IPR006520; Pro_4_hyd_alph.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF031171; 2OG-FeII_Oxy; 1.
DR SMART; SM00702; F4Hc; 1.
DR PROSITE; PS00014; ER_TARGET; 1.
SQ SEQUENCE 739 AA; 83592 MW; 64B7AE68E80B68EA CRC64;

Query Match 87.9%; Score 3403.5; DB 11; Length 739;
Best Local Similarity 87.4%; Pred No. 1.9e-247;
Matches 646; Conservative 34; Mismatches 56; Indels 3; Gaps 2;

Qy 1 MAVRALKLLTLLAVAAAS-QAEVESEAGCMVTPDLLFAEGTAAYAGDWPGLVLSME 59
Db 1 MAVSERLLAAMLAIAAAALRAVAASEPGWDVAAPDLLYAEGTAAYAGDWPGLVLSME 60
Qy 60 RALRSRALRALRCCTCCGADPPWELDPDWSPSP--ACASGAGALRDLSPFGGLIRRA 117
Db 61 RALRSRALRALRCRTRCATLFWAPDLDLGDPDLSLQDPGAAALHDLRFFGVAIRRA 120

```

```

Qy 118 ACIRCLGPPAAHSISEMELFEFRKSPYNYLOVAYFKINKLEKAVAAAHFTFFVGNPEHM 177
Db 121 ACIRCLGPPSAHLSSELDLEFNKRSYNYLOVAYFKINKLEKAVAAAHFTFFVGNPEHM 180
Qy 178 EMQNQLDDYQTMASGVKEADFDKLETPQHMQBFRLGVRLYSEBQPOEAVPHLEAALQEVFV 237
Db 181 EMQNQLDDYQTMASGVKEADFDLEAKPHMHEPRLGVRLYSEBQPOEAVPHLEAALQEVFV 240
Qy 238 AVEECRALCEGYDYGNYLYNADLFOAITDHYIQLVNCNKQNCVTELASHPSREKPF 297
Db 241 ADEECRALCEGYDYGNYLYNADLFOAITDHYVQVQVNCNKQNCVTELASHPSREKPF 300
Qy 298 DFLPSHYNLQFAYNIGNYTOAGCACTYLLFFPNDEVMNQNLAYYAAMLGEEHTRISIG 357
Db 301 DFLPSHYNLQFAYNIGNYTOAIECACTYLLFFPNDEVMNQNLAYYTAMLGEEASIS 360
Qy 358 PRESAREYRQSRSLLEKELLFFAYDVGIGPFVDPDSWTPPEVPIKGLQEKOKSERETAVRI 417
Db 361 PRENAEYRRRLLEKELLFFAYDVGIGPFVDPDSWTPPEVPIKRLQEKOKSERETAVRI 420
Qy 418 SOEIGNLMKEIETLVEEKTESLVSRLTREGGPLYEGISLTMNSKLLNGYQRVVMGDV 477
Db 421 SOEIGNLMKEIETLVEEKTESLVSRLTREGGPLYEGISLTMNSKLLNGYQRVVMGDV 480
Qy 478 ISDHCEQELQRLTNVAATSGDGYRGQTSPTPNKFKYGVTVFKALKLQEGKVPLOSAHL 537
Db 481 ISDDECEQELQRLTNAAATSGDGYRGQTSPTPNKFKYGVTVFKALKLQEGKVPLOSAHM 540
Qy 538 YYNVTEKVRIRIMESYFRDLTPLYFSYSHLVCRTAIEVQAEKODSDHPVHVDNCILNAET 597
Db 541 YYNVTEKVRIMESYFRDLTPLYFSYSHLVCRTAIEESQAEKODSSHPVHVDNCILNAEA 600
Qy 598 LVCVKEPPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAEVQPCGRAVGFSSGTEN 657
Db 601 LNCIKEPPAYTFRDYSAILYLNGDFDGNFYFTLDAKTVAEVQPCGRAVGFSSGTEN 660
Qy 658 PHGVKAVTRGORCAIALWFTLDRHRSERDRVQADDLVKMLSPSEMDLSQBPDLDAQGP 717
Db 661 PHGVKAVTRGORCAIALWFTLDRHRSERDRVQADDLVKMLSPSEMDLSQBPDLPGQOQS 720
Qy 718 PEPAOESSLSGSSEKPKDEL 736
Db 721 PEFGEESLSDRGSLHKDEL 739

RESULT 6
Q8BSV2 PRELIMINARY; PRT; 738 AA.
AC Q8BSV2
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Leprecaan 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Pituitary;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK030436; BAC26962.1; -.
DR PIR; PT0551; PT0551.
DR GO; GO:0016706; F:oxidoreductase activity, acting on paired d. . .; IEA.
DR GO; GO:0019538; P:protein metabolism; IEA.
DR InterPro; IPR005123; 2OG-FeII_Oxy.
DR InterPro; IPR000886; ER_target_S.
DR InterPro; IPR006620; Pro_4_hyd_alph.

```



Query Match	87.3%	Score 3379	DB 11	Length 738
Best Local Similarity	87.1%	Pred. No. 1.3e-245		
Matches 644	Conservative 34	Mismatches 57	Indels 4	Gaps 3
DR InterPro: IPR008941; TPR-like.				
DR Pfam: PF03171; 2OG-Feril Oxy; 1.				
DR SMART: SM00702; P4HC; 1.				
DR PROSITE: PS00014; ER TARGET; 1.				
SQ SEQUENCE 738 AA; 83528 MW; 75967DE318F55F4A CRC64;				
QY 1 MAVRALKLLTLLAVAAAS-QAEVSEAGMGWVTPDLLFAEGTAAYARGDPGWGLSME 59				
DB 1 MAVERRLLAAMLAVAANAALRVAASEPGWDVAAPDLLLYAEGTAAYSRGDPGWGLNME 60				
QY 60 RALRSRAALRALRURCTQCAADFPWELDPWSP--AQASGAGALRDLSPFGGLLRRA 117				
DB 61 RALRSRAALRALRURCTRCATELFWAPDLDLGDPDPSLSDQPGAAALHDLFFGAVLRRA 120				
QY 118 ACLRCLGPPAAHSLSEMELEFKRSPYNLYQVAYFKINKLEKAVAAAHFFVGNPEHM 177				
DB 121 ACLRCLGPPSAHLLSELDLEFKRSPYNLYQVAYFKINKLEKAVAAAHFFVGNPEHM 180				
QY 178 EMQNLDYYQTMGSGYEADFKDLQTPHQMFRLGVRLYSEQQAEPVPHLEALQRYFV 237				
DB 181 EMQNLDYYQTMGSGYEADFDLEAKPHMHSEFRILGVRLYSEKQAEPVPHLEALQRYFV 240				
QY 238 AYEBCRALCEGPDYDGVNLYEYNADLFOALTTHYIOVLNCKQNCVTELASHPSREKPF 297				
DB 241 ADEBCRALCEGPDYDGVNLYDYSADLFQALTTHYIOVLNCKQNCVTELASHPSREKPF 300				
QY 298 DFLPSHYNLYQFAFYNIQNYTQAQCEAKTYLLFFPNDEVMNQNLAYYAAMLGEBHRSIG 357				
DB 301 DFLPSHYNLYQFAFYNIQNYTQAQCEAKTYLLFFPNDEVMHQNLAAYTAMLGEEEAASSIS 360				
QY 358 PRESAYEYRQSRLLSEKELLFPAYDVFGIPFVDPSWTPPEEVIKPELQEKQSERETAARI 417				
DB 361 PRENAEYRRRESLLEKELLFPAYDIFGIPFVDPSWTPPEEVIKPELQEKQSERETAAR- 419				
QY 418 SOETGNLMKETTLVEEKTESLVSRLTRREGPLLLEGISLTWNKLLNGYQRVVMDGV 477				
DB 420 SOETGNLMKETTLVEEKTESLVSRLTRREGPLLLEGISLTWNKLVNGSQRVVMDGV 479				
QY 478 ISDHECQELQRLTNVAATSGDGYRGQTSPHTPNEKFGYGVTVFKALKLQEGKVPLOSAHL 537				
DB 480 ISDDECQELQRLTNAAATSGDGYRGQTSPHTPNEKFGYGVTVFKALKLQEGKVPLOSAHM 539				
QY 538 YYNVTEKVRRTWESYFRDLTPLYSYSHLVCRTALEEVQAEKDDSHPVHVDNCLINAEI 597				
DB 540 YYNVTEKVRRTWESYFRDLTPLYSYSHLVCRTALEESQAEKDDSHPVHVDNCLINAEA 599				
QY 598 LVCVKEPPAYFRDYSAILYLNGDDPGGNFYFTELDAKTVTAEVQPCQRAVGFSSGTE 657				
DB 600 LMCVKEPPAYFRDYSAILYLNGDDPGGNFYFTELDAKTVTAEVQPCQRAVGFSSGTE 659				
QY 658 PHGVKAVTRGORCAIALNFTLDPHSEDRVQADDLVKMLSPSEEMDLSSQEQPLDAQGP 717				
DB 660 PHGVKAVTRGORCAIALNFTLDPHSEDRVQADDLVKMLSPSEEVLDLPQEQPLDQOQS 719				
QY 718 PEPAQESLSGSESKPKDEL 736				
DB 720 PEPGEESLSDRGSLHKDEL 738				
RESULT 7				
ID Q9RLJ8	PRELIMINARY;	PRT;	728 AA.	
AC Q9RLJ8;				
DT 01-MAY-2000 (TREMBLrel. 13, Created)				
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)				
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)				
DE Leprecan.				
OS Rattus norvegicus (Rat).				

PRECISE.T 7

RESULTS /					
Q9RLJ8					
ID	Q9RLJ8	PRELIMINARY;	PRT; 728 AA.		
AC	Q9RLJ8;				
DT	01-MAY-2000	(TREMBlrel. 13, Created)			
DT	01-MAY-2000	(TREMBlrel. 13, Last sequence update)			
DE	01-OCT-2003	(TREMBlrel. 25, Last annotation update)			
DT					
LE	Leprrecan.				
GN	LEPRel.				
OS	Rattus norvegicus (Rat).				
<hr/>					
541	MESYFELDTELYFSYSHLVCRTAIBESQAERKDSHPVHVNDICILNAESLYVICKEPPAYT	Db			600
609	FRIYSAILYLNGDFGCGNEFFFTTELDAKTVTAVSQCGRAVCSFGSTGNPHGVKAVTRGQ	Qy			668
601	FRIYSAILYLNGDFDGNGFFFTTELDAKTVTAVSQCGRAVCSFGSTGNPHGVKAVTRGQ	Db			660
669	RCALIAWFTLDPRHSERDRVQADDLVKMLFSPSEMDLSQEQLDLQQCFPPPAQSLSGS	Qy			728
661	RCALIAWFTLDPRHSERDRVQADDLVKMLFSPSEVDLPQEQLPQQCSFKPKGESLSDR	Db			720

```
QY 729 ESKPKDEL 736
DB 721 ESQPKDEL 728

RESULT 8
Q9QZT7
ID ID Q9QZT7 PRELIMINARY; PRT; 747 AA.
AC Q9QZT7;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Growth suppressor 1L.
GN LEPRE1 OR GROS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD1 ICR; TISSUE=Fibroblast;
RX MEDLINE=20406537; PubMed=10951563;
RA Kaul S.C., Sugihara T., Yoshida A., Nomura H., Wadhwa R.;
RT "Gros1, a potential growth suppressor on chromosome 1: its identity to
RL basement membrane-associated proteoglycan, leprecan.";
DR EMEL; AF165163; AAF04806.1; -.
DR MGD; MGI:1888921; Leprel.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0016049; P:cell growth; IDA.
DR InterPro; IPR005123; 2OG-FeII_Oxy.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; 2OG-FeII_Oxy; 1.
DR SMART; SM00702; P4Hc; 1.
SQ SEQUENCE 747 AA; 84811 MW; EBBAB2A62E824123 CRC64;

Query Match 85.7%; Score 3316; DB 11; Length 747;
Best Local Similarity 88.2%; Pred. No. 7.4e-241;
Matches 623; Conservative 33; Mismatches 48; Indels 2; Gaps 1;

QY 25 ESEAGWGWTPDLFLFAGTAAAYAGDWPVGLSMERALSRALRALRLRCRQCQAADFP 84
DB 25 ESEPGWVAAPDLLLYAGTAAYSRRDWPVGLVNERALSRALRALRLRCRQCATELP 84
QY 85 WELDPDWSPP--AQASGAGALRDLSPFGGLLRRAACLRCLGPPAAHLSIEEMELEPRK 142
DB 85 WAPDLGLDPSPSLSDQDCAALHDLRFPFAGVLRRAACLRCLGPPSAHLLSELDLEFNK 144
QY 143 RSPNYLQVAYFKINKLEKAAVAHAHFFVGNPHMEMQNLDYQYTMGSKVKEADPKDLFT 202
DB 145 RSPNYLQVAYFKINKLEKAAVAHAHFFVGNPHMEMQNLDYQYTMGSKVKEADPKDLFT 204
QY 203 QPHMQEPLGLVRLYSEBPQEAIPHLEAALQYFVAVEECRALCEGPDYDGVNLYEYNA 262
DB 205 KPHMHFPLGLVRLYSEBKPQEAIPHLEAALQYFVADEECRALCEGPDYDGVNLYDYS 264
QY 263 DLQAITDHYIOLVNLCKQNCVTELASHPSREKPFEDPLSHYNYLQFAYNYGNITQAGE 322
DB 265 DLQAITDHYVQVNLCKQNCVTELASHPSREKPFEDPLSHYNYLQFAYNYGNITQAGE 324
QY 323 CAKTYLLFFPNDEVMQNLAYYAAMLCEETRTSIGPREGSAKEYRQSRLLKXELLFPAYDV 382
DB 325 CAKTYLLFFPNDEVMQNLAYYAAMLCEETRTSIGPREGSAKEYRQSRLLKXELLFPAYDI 384
QY 383 FGIPFPDPSWTPEBVIIPKRLQKQKSERETAVRISQEIIGNLMKEITETLVEEKTESLDV 442
DB 385 FGIPFPDPSWTPEBVIIPKRLQKQKSERETAVRISQEIIGNLMKEITETLVEEKTESLDV 444
QY 443 SRLTREGGPLLYEGISLTWNSKVLNGSQRVVMDGVISDDECEQLQRLTNAATSGDGYRG 504
DB 503 QTSPTNEKEFYGVTVFKALKLQGGKVPLOSAGHLYYNNVTEKVRIMESYRLDTPLYFS 562
DB 505 QTSPTNEKEFYGVTVFKALKLQGGKVPLOSAGHLYYNNVTEKVRIMESYRLDTPLYFS 564
QY 563 YSHLVCRTAIEBQVQAEKDDSHPVHVDNCILNAETLVCVKEPPAYTFRDYGAILYNGDF 622
DB 565 YSHFVCRTAIEBQVQAEKDDSHPVHVDNCILNAEAFMCIEPPAYTFRDYGAILYNGDF 624
QY 623 DCGNFYFTELDAKTVAEVOQCGRAVGFSSGTENPHGVKAVTRGQRCALALNTLDRH 682
DB 625 DCGNFYFTELDAKTVAEVOQCGRAVGFSSGTENPHGVKAVTRGQRCALALNTLDRH 684
QY 683 SERDRVQADDLVKMLFSPPEMDLSQEQPLDAQQPPPEPAQESLSGS 728
DB 685 SERDRVQADDLVKMLFSPPEMDLSQEQPLDAQQPPPEPAQESLSGS 730

RESULT 9
Q9CWK5
ID ID Q9CWK5 PRELIMINARY; PRT; 742 AA.
AC Q9CWK5;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE 2410024C1SR1K protein.
GN LEPRE1 OR GROS1 OR 2410024C1SR1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryonic stem cells;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Iehi Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schirni L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK010578; BAB27041.1; -.
DR MGD; MGI:1888921; Leprel.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0016049; P:cell growth; IDA.
DR InterPro; IPR005123; 2OG-FeII_Oxy.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; 2OG-FeII_Oxy; 1.
DR SMART; SM00702; P4Hc; 1.
SQ SEQUENCE 742 AA; 84086 MW; CFC44D34D96ACEA7 CRC64;

Query Match 85.4%; Score 3305; DB 11; Length 742;
Best Local Similarity 86.0%; Pred. No. 4.9e-240;
Matches 629; Conservative 37; Mismatches 57; Indels 8; Gaps 3;

QY 1 MAVFALKLLTLLAVAAAS-QAEVSEFAGWGVTPDLLFAEGTAAVARGWFGVLSME 59
```

Db 1 MAVSERLLAAMLA... 60  
QY 60 RALRSAAALRALRLCRTCAADFPWELDPDWSPP--AAAGAGALRLDLSPFGGLLRA 117  
Db 61 RALRSAAALRALRLCRTCAATLPWAPDLDLGDPFSLQSDPFAAALHDLRFFGAVLRA 120  
QY 118 ACLRRCLGPPAAHSLSEEMLEFRKSPYNYLQVAFKINKLEKAVAAAHTEFFVGNPEHM 177  
Db 121 ACLRRCLGPPSAHLLSEELDLSPFNKSPYNYLQVAFKINKLEKAVAAAHTEFFVGNPEHM 180  
QY 178 EMQNLDDYQYQVSGVKEADFKOLETOPHMOERFLGRLVLYSEPOQAVPHLEAALQEVFV 237  
Db 181 EMQNLDDYQVSGVKEADFRDLKAPHEPRFLGRLVLYSEPOQAVPHLEAALQEVFV 240  
QY 238 AVEECALCEGPDYDGYNYLYENADLFOAITDHYIQLVNLCKQNCVTELASHPSREKPE 297  
Db 241 ADEECALCEGPDYDGYNYLYENADLFOAITDHYIQLVNLCKQNCVTELASHPSREKPE 300  
QY 298 DFLPSHYNLYQFAYYNYGNYTQAGECAKTYLLFFPNDEVMQNQLAYYAAAMLGEEHTRSG 357  
Db 301 DFLPSHYNLYQFAYYNYGNYTQAGECAKTYLLFFPNDEVMQNQLAYYAAAMLGEEHTRSG 360  
QY 358 PRESAREYRRLSLEKELLFFAYDVFEGIPFVDPDSWTPEVTPKRLQEKOKSERETAVRI 417  
Db 361 PRENABEYRRLSLEKELLFFAYDVFEGIPFVDPDSWTPEVTPKRLQEKOKSERETAVRI 420  
QY 418 SQEIGNLMKEIETLVBEKTESLDVSRKLTREGGPLYEGISLTMSKLVNGYQVVMDCV 477  
Db 421 SQEIGNLMKEIETLVBEKTESLDVSRKLTREGGPLYEGISLTMSKLVNGYQVVMDCV 480  
QY 478 ISDHCEQELQRLNVAATSGDGYRGTSPTHPNEKFGYGVTVFVKALKLGQEGKVPLOSALH 537  
Db 481 ISDHCEQELQRLNVAATSGDGYRGTSPTHPNEKFGYGVTVFVKALKLGQEGKVPLOSALH 540  
QY 538 YNNVTEKVRIMESYRLDPLFYSHLVCRTAIEEVOAERKDDSHPVHVDNCLNAET 597  
Db 541 YNNVTEKVRIMESYRLDPLFYSHLVCRTAIEEVOAERKDDSHPVHVDNCLNAET 600  
QY 598 LVCVKEPPATFRDYSAIYLNGDFDGNFYFTELDPAKTVTAEVQPCGRVAFSSGTEN 657  
Db 601 LMCIEKPPATFRDYSAIYLNGDFDGNFYFTELDPAKTVTAEVQPCGRVAFSSGTEN 660  
QY 658 PHGVKAVTRQORCAIALWFTLDRHSERDRVQADLVKMLFSPPEMDLSQEQPLDAQGP 717  
Db 661 PHGVKAVTRQORCAIALWFTLDRHSERDRVQADLVKMLFSPPEMDLSQEQPLDAQGP 718  
QY 718 PEPAOESLSGS 728  
Db 719 ---ARQSLKS 726

RESULT 10  
Q9QZT6 PRELIMINARY; PRT; 542 AA.  
ID Q9QZT6  
AC Q9QZT6  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Growth suppressor 1S.  
GN LEPREL OR GROS1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CD1 ICR;  
RX MEDLINE=20406537; PubMed=10951563;  
RA Kaul S.C., Sugihara T., Yoshida A., Nomura H., Wadhwa R.;  
RT "Gros1, a potential growth suppressor on chromosome 1: its identity to  
RT basement membrane-associated proteoglycan, leprecan."  
RL Oncogene 19:3576-3583 (2000).

DR EMBL; AF165164; AAF04807.1; -.  
DR MGD; MGI:188921; Leprel.  
DR GO; GO:0005737; C:cytoplasm; IDA.  
DR GO; GO:0005634; C:nucleus; IDA.  
DR GO; GO:0016049; P:cell growth; IDA.  
DR InterPro; IPR008941; TPR-like.  
SQ SEQUENCE 542 AA; 61599 MW; 0A211411F71DF168 CRC64;  
Query Match 61.6%; Score 2384; DB 11; Length 542;  
Best Local Similarity 87.9%; Pred. No. 7.2e-171;  
Matches 451; Conservative 25; Mismatches 35; Indels 2; Gaps 1;  
QY 25 ESEAGWMTVDLLPAEGTAAYAGDWPVGLVSMERALSRAALRALRLCRTCAADFP 84  
Db 25 ESEAGWMTVDLLPAEGTAAYAGDWPVGLVSMERALSRAALRALRLCRTCAADFP 84  
QY 85 WELDPDWSPP--AAAGAGALRLDLSPFGGLLRAAALRLCLGPPAAHSLSEEMLEFRK 142  
Db 85 WAPDLDLGDPFSLQSDPFAAALHDLRFFGAVLRAAALRLCLGPPSAHLLSEELDFENK 144  
QY 143 RSPYNYLQVAFKINKLEKAVAAAHTEFFVGNPEHMEMQNLDYQYQVSGVKEADFKLET 202  
Db 145 RSPYNYLQVAFKINKLEKAVAAAHTEFFVGNPEHMEMQNLDYQYQVSGVKEADFKLET 204  
QY 203 QPHMOERFLGRLVLYSEPOQAVPHLEAALQEVFVAYVEECRALCEGPDYDGYNYLYENA 262  
Db 205 KPHHEFFLGRLVLYSEPOQAVPHLEAALQEVFVAYVEECRALCEGPDYDGYNYLYENA 264  
QY 263 DLFQAITDHYIQLVNLCKQNCVTELASHPSREKPEDEFLPSHYNLYQFAYYNYGNYTQAGE 322  
Db 265 DLFQAITDHYIQLVNLCKQNCVTELASHPSREKPEDEFLPSHYNLYQFAYYNYGNYTQAGE 324  
QY 323 CAKTYLLFFPNDEVMQNQLAYYAAAMLGEEHTRSGPRESAREYRRLSLEKELLFFAYDI 382  
Db 325 CAKTYLLFFPNDEVMQNQLAYYAAAMLGEEHTRSGPRESAREYRRLSLEKELLFFAYDI 384  
QY 383 FGIPFVDPDSWTPEVTPKRLQEKOKSERETAVRISQIGNLMKEIETLVBEKTESLDV 442  
Db 385 FGIPFVDPDSWTPEVTPKRLQEKOKSERETAVRISQIGNLMKEIETLVBEKTESLDV 444  
QY 443 SELLTREGGPLYEGISLTMSKLVNGYQVVMDCVVISDHCEQELQRLNVAATSGDGYRG 502  
Db 445 SELLTREGGPLYEGISLTMSKLVNGYQVVMDCVVISDHCEQELQRLNVAATSGDGYRG 504  
QY 503 QTSPTHPNEKFGYGVTVFVKALKLGQEGKVPLOSA 535  
Db 505 QTSPTHPNEKFGYGVTVFVKALKLGQEGKVPLOSA 537

RESULT 11  
Q96BR8 PRELIMINARY; PRT; 390 AA.  
ID Q96BR8  
AC Q96BR8  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Growth suppressor 1 (Leucine proline-enriched proteoglycan (leprecan)  
DE 1).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Ductenium;  
RA Strausberg R.;  
RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Kalline N., Chen X., Rolfs A., Halleck A., Hines L., Eisenstein S.,  
RA Koundinya M., Raphael J., Moreira D., Kelley T., LaBaer J., Lin Y.,  
RA Phelan M., Farmer A.;  
RT "Cloning of human full-length CDSs in BD Creator(TM) System Donor

RT vector.";  
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC015309; AAH15309.1; -.  
 DR EMBL; BT007039; AAP35688.1; -.  
 DR GO; GO:0016706; P:oxidoreductase activity, acting on paired d. . .; IEA.  
 DR GO; GO:0019538; P:protein metabolism; IEA.  
 DR InterPro; IPR005123; 2OG-FeII\_Oxy.  
 DR InterPro; IPR000886; ER\_target\_S.  
 DR Pfam; PF03171; 2OG-FeII\_Oxy; 1.  
 DR SMART; SM00702; P4HC; 1.  
 DR PROSITE; PS00014; ER\_TARGET; 1.  
 SQ SEQUENCE 390 AA; 44077 MW; 61C9F5983CF637FB CRC64;

Query Match 52.3%; Score 2024; DB 4; Length 390;  
 Best Local Similarity 99.7%; Pred. No. 5.7e-144;  
 Matches 389; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 347 MLGEHTRSIGPRESAKYRQSLLEKELLPFAVDVFGIPFVDPDPSWTPEEVIKRLQEK 406  
 DB 1 MLGEHTRSIGPRESAKYRQSLLEKELLPFAVDVFGIPFVDPDPSWTPEEVIKRLQEK 60

QY 407 QKSERETAVRISQEIENLMKEIETLVEKTKESLDVSRLTREGGPLYEGISLTMSKLL 466  
 DB 61 QKSERETAVRISQEIENLMKEIETLVEKTKESLDVSRLTREGGPLYEGISLTMSKLL 120

QY 467 NGYQVRVMDGVISDHECOELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLQ 526  
 DB 121 NGYQVRVMDGVISDHECOELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLQ 180

QY 527 EGVKPLQSAHLVYNTKVRIMESYFRDLTPLYFSYSHLVCRITAIEEVOAERKDDSHPV 586  
 DB 181 EGVKPLQSAHLVYNTKVRIMESYFRDLTPLYFSYSHLVCRITAIEEVOAERKDDSHPV 240

QY 587 HVDNCILNAETLVCKEPPATFRDYSAIYLNGDFDGGNFYFTELDAAKTVTAEVQPOCG 646  
 DB 241 HVDNCILNAETLVCKEPPATFRDYSAIYLNGDFDGGNFYFTELDAAKTVTAEVQPOCG 300

QY 647 RAVGFSSGTENPHGVKAVTRGQRCALWFTLDPHRSERDRVQADDLVKMLFSPFEMDLS 706  
 DB 301 RAVGFSSGTENPHGVKAVTRGQRCALWFTLDPHRSERDRVQADDLVKMLFSPFEMDLS 360

QY 707 QEQLDAQCGPPPEPAQESLSGSSEKPKDEL 736  
 DB 361 QEQLDAQCGPPPEPAQESLSGSSEKPKDEL 390

RESULT 12  
 Q9H6K3 PRELIMINARY; PRT; 390 AA.  
 AC Q9H6K3;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein FLJ22188.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,  
 RA Suzuki Y., Iobayashi M., Nishi T., Shibahara T., Tanaka T.,  
 RA Nakamura Y., Ohtsuka T., Sugano S.;  
 RT "NEDO human cDNA sequencing project";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK025841; BAB15256.1; -.  
 DR GO; GO:0016706; F:oxidoreductase activity, acting on paired d. . .; IEA.  
 DR GO; GO:0019538; P:protein metabolism; IEA.  
 DR InterPro; IPR005123; 2OG-FeII\_Oxy.  
 DR InterPro; IPR000886; ER\_target\_S.  
 DR Pfam; PF03171; 2OG-FeII\_Oxy; 1.

DR SMART; SM00702; P4HC; 1.  
 DR PROSITE; PS00014; ER\_TARGET; 1.  
 KW Hypothetical protein\_  
 SQ SEQUENCE 390 AA; 44093 MW; 61D29959310E677FB CRC64;

Query Match 52.0%; Score 2014; DB 4; Length 390;  
 Best Local Similarity 99.5%; Pred. No. 3.2e-143;  
 Matches 388; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 347 MLGEHTRSIGPRESAKYRQSLLEKELLPFAVDVFGIPFVDPDPSWTPEEVIKRLQEK 406  
 DB 1 MLGEHTRSIGPRESAKYRQSLLEKELLPFAVDVFGIPFVDPDPSWTPEEVIKRLQEK 60

QY 407 QKSERETAVRISQEIENLMKEIETLVEKTKESLDVSRLTREGGPLYEGISLTMSKLL 466  
 DB 61 QKSERETAVRISQEIENLMKEIETLVEKTKESLDVSRLTREGGPLYEGISLTMSKLL 120

QY 467 NGYQVRVMDGVISDHECOELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLQ 526  
 DB 121 NGYQVRVMDGVISDHECOELQRLTNVAATSGDGYRGQTSPTNEKFGYGVTVFKALKLQ 180

QY 527 EGVKPLQSAHLVYNTKVRIMESYFRDLTPLYFSYSHLVCRITAIEEVOAERKDDSHPV 586  
 DB 181 EGVKPLQSAHLVYNTKVRIMESYFRDLTPLYFSYSHLVCRITAIEEVOAERKDDSHPV 240

QY 587 HVDNCILNAETLVCKEPPATFRDYSAIYLNGDFDGGNFYFTELDAAKTVTAEVQPOCG 646  
 DB 241 HVDNCILNAETLVCKEPPATFRDYSAIYLNGDFDGGNFYFTELDAAKTVTAEVQPOCG 300

QY 647 RAVGFSSGTENPHGVKAVTRGQRCALWFTLDPHRSERDRVQADDLVKMLFSPFEMDLS 706  
 DB 301 RAVGFSSGTENPHGVKAVTRGQRCALWFTLDPHRSERDRVQADDLVKMLFSPFEMDLS 360

QY 707 QEQLDAQCGPPPEPAQESLSGSSEKPKDEL 736  
 DB 361 QEQLDAQCGPPPEPAQESLSGSSEKPKDEL 390

RESULT 13  
 Q9HC87 PRELIMINARY; PRT; 363 AA.  
 AC Q9HC87;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE GROS1-S protein.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA TISSUE-Testis;  
 RC MEDLINE=20406537; PubMed=10951563;  
 RA Kaul S.C., Sugihara T., Yoshida A., Nomura H., Wadhwa R.;  
 RT "Gros1, a potential growth suppressor on chromosome 1: its identity to  
 RT basement membrane-associated proteoglycan, leprecan.";  
 RL Oncogene 19:3576-3583(2000).  
 DR EMBL; AF097431; AAG31018.1; -.  
 DR GO; GO:0008285; P:negative regulation of cell proliferation; NAS.  
 DR InterPro; IPR008941; TPR-like.  
 DR SEQUENCE 363 AA; 41158 MW; 5B1B66B38679E76A CRC64;

Query Match 49.4%; Score 1911; DB 4; Length 363;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-135;  
 Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NAVFALKLLTLLAVAAASQAEVSEAGAGMTPTDLLFAEGTAAAYARGDPGVLSMER 60  
 DB 1 NAVFALKLLTLLAVAAASQAEVSEAGAGMTPTDLLFAEGTAAAYARGDPGVLSMER 60

QY 61 ALRGRALRLRCRTCCAADPFWELDPDWSPPAQAGALRDLSPFGGLLRACL 120

```
Db 61 ALKRAALRALRLCRCTOCAAFFWELDPDWSPSPAQAGAGALRDLSPFGGLLRAACL 120
QY 121 RRCGLGPPAAHSLSEEMELFRKSPNYLQVAYFKINKLEKAAVAHAHTFFVGNPEHMEMQ 180
Db 121 RRCGLGPPAAHSLSEEMELFRKSPNYLQVAYFKINKLEKAAVAHAHTFFVGNPEHMEMQ 180
QY 181 QNLDDYQTMGSKVKEADFQKLETOHPHQEPRGLGLVRLYSEPOQEAHPHLEAALQEFVAYE 240
Db 181 QNLDDYQTMGSKVKEADFQKLETOHPHQEPRGLGLVRLYSEPOQEAHPHLEAALQEFVAYE 240
QY 241 ECRALCEGPDYDGYNLYEVNADLFOAITDHYIQLVNLCKQNCVTELASHPSREKPEDEFL 300
Db 241 ECRALCEGPDYDGYNLYEVNADLFOAITDHYIQLVNLCKQNCVTELASHPSREKPEDEFL 300
QY 301 PSHYNYLQFAYNIGNYTOAGECAKTYLLFFPNDEVMNQNLAYYAAMLGEHTRSGIPRE 360
Db 301 PSHYNYLQFAYNIGNYTOAGECAKTYLLFFPNDEVMNQNLAYYAAMLGEHTRSGIPRE 360

RESULT 14
Q81VL5 PRELIMINARY; PRT; 708 AA.
ID AC Q81VL5
AC Q81VL5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Myxoid liposarcoma associated protein 4.
GN MLAT4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Thelin-Jarnum S., Aman P.;
RT "The novel gene Myxoid liposarcoma Associated Transcript 4 (MLAT4) belongs to the Gro1/leprecan proteoglycan family.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ430351; CA223029.2;
DR GO; GO:0016706; F:oxidoreductase activity, acting on paired d. . . ; IEA.
DR GO; GO:0019538; P:protein metabolism; IEA.
DR InterPro; IPR005123; 2OG-Fell_Oxy.
DR InterPro; IPR000886; ER target S.
DR InterPro; IPR006620; PrO 4_hyd_alph.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; 2OG-Fell_Oxy; 1.
DR SMART; SM00702; P4HC; 1.
DR PROSITE; PS00014; ER TARGET; 1.
SQ SEQUENCE 708 AA; 80984 MW; B9E680C90D607291 CRC64;

Query Match 45.1%; Score 1747; DB 4; Length 708;
Best Local Similarity 49.2%; Pred. No. 1.1e-122;
Matches 355; Conservative 103; Mismatches 211; Indels 52; Gaps 12;

QY 23 ESEAGWGMVTP-DLLFAECTAAYARGDVGWVLSMERALRSRALRLRCRTQCAA 81
Db 33 ELELEP-GPIQPTPELLYASGAAAYSGDYERAVRDLEAALRSRRLREINTRCARCAA 90
QY 82 DFPWELDPDWSPAQAGALRDLSPFGGLLRAACLRC-----LGPPAA-HSLSEEM 136
Db 91 RHP-----LPPPPPGGPGA--ELPLRLLGRARCRCCTQRLOGGPRASRHRVSEDV 141
QY 137 ELEFRKSPNYLQVAYFKINKLEKAAVAHAHTFFVGNPEHMEMQNLDYQTMGSKVKEAD 196
Db 142 RSDQRQRPVNYLQFAYIKLQLEKAVEAHTFFVGNPEHMEMQNENTYRATAGVEALQ 201
QY 197 FKDLTQPHMQEPRGLVLYSEPOQEAHPHLEAALQEFVAYECCRALCEGPDYDGYN 256
Db 202 LVDRKAPKPMESYNAGVQKYEADDPEMAIRHFEQALREYFVEDTECRCLCEGPRQFBEYE 261
QY 257 YLEVNADLFOAITDHYIQLVNLCKQNCVTELASHPSREKPEDEFLPSHYNYLQFAYNIGN 316
Db 262 YLGYKAGLYEAIADHYMQVLCQHECVRELATRCGLSPFLENFLPHDYDLQFAYYRVE 321
```

```
QY 317 YTOAGECAKTYLLFFPNDEVMNQNLAYYAAMLGEE-HTRSIGPRESAKKEYRQSRLEKEL 375
Db 322 YVKALEKAKAYLLCHPDDDEDVLDNDVDYSESLDDSDIDPASTEAREDLTMFVKRHKLSSEL 381
QY 376 LFFAYDVGIPFVDPDSWTPPEVVPKRLQEKQKSERETAFAVRSQIEIGNLMKEIETLVEEK 435
Db 382 IKSAAGLGFYTFPNYWI-----RYGRQDENRVPGVNVEGAHVHGFSGMK 429
QY 436 TKESLDVSRLLTFREGPLLYEGISLTMSKLNGLNGVQVMDGVI SDHECOELQRLTNVAAT 495
Db 430 -KLSPKIDREDLREGPLLYENITFYVNSEQNGTQRVLLDNLVSEEQCRELHVSASGIML 488
QY 496 SDGVRGQTSPTNPKXFTGVTVFKALKLGOEGVPLOSALHYXNVTKVRIMESYFRL 555
Db 489 VDGVRGKTSPTNPKXFTGVTVFKALKLGOEGVRPLKSARLFYDISEKARRIVESYFML 548
QY 556 DTPLYFSYSHLVCRTAIEEVAERKDDSHPVHVNCLNAETLVCVKPEPPATYTRDYSAI 615
Db 549 NSTLYFSYTHMVCRTALSGQQRNDLSHPHIAHNCILDPANECWKEPPATYTRDYSAI 608
QY 616 LYINGDFDGNFYTFELDAKTTVAEVQPCGRAVGFSSGTENPHGVKAVTRGORCAIALW 675
Db 609 LYMDDDFEGEFIFTEMDAKTVTASIKPKCGRMISFSSGGENPHGVKAVTRGORCAIALW 668
QY 676 FTLDPRHSERDRVQADLLVXMLFSPPEMDLSQEQPLDAQQGPPEPAQESLSGSESKPDE 735
Db 669 FTLDPLYRELERIQADEVIAI-----LDQE-----QQGKHE-----LNINPKDE 707
QY 736 L 736
Db 708 L 708

RESULT 15
Q8CG71 PRELIMINARY; PRT; 703 AA.
ID AC Q8CG71
AC Q8CG71;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Myxoid liposarcoma associated protein 4.
GN AWS53532 OR MLAT4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Thelin-Jarnum S., Aman P.;
RT "The novel gene Myxoid liposarcoma Associated Transcript 4 (MLAT4) belongs to the Gro1/leprecan proteoglycan family.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ430350; CAD23038.1;
DR MGD; MGI:2146663; AWS53532.
DR GO; GO:0016706; F:oxidoreductase activity, acting on paired d. . . ; IEA.
DR GO; GO:0019538; P:protein metabolism; IEA.
DR InterPro; IPR005123; 2OG-Fell_Oxy.
DR InterPro; IPR000886; ER target S.
DR InterPro; IPR006620; PrO 4_hyd_alph.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF03171; 2OG-Fell_Oxy; 1.
DR SMART; SM00702; P4HC; 1.
DR PROSITE; PS00014; ER TARGET; 1.
SQ SEQUENCE 703 AA; 80154 MW; 1F42F9B9938573E4 CRC64;

Query Match 44.3%; Score 1715.5; DB 11; Length 703;
Best Local Similarity 47.6%; Pred. No. 2.5e-120;
Matches 342; Conservative 105; Mismatches 219; Indels 53; Gaps 8;

QY 25 ESEAGWGMVTP-DLLFAECTAAYARGDVGWVLSMERALRSRALRLRCRTQCAA 83
```



Result No.	Score	Query Match	Length	DB ID	Description
1	100	100	100	1	...
2	95	95	95	2	...
3	90	90	90	3	...
4	85	85	85	4	...
5	80	80	80	5	...
6	75	75	75	6	...
7	70	70	70	7	...
8	65	65	65	8	...
9	60	60	60	9	...
10	55	55	55	10	...
11	50	50	50	11	...
12	45	45	45	12	...
13	40	40	40	13	...
14	35	35	35	14	...
15	30	30	30	15	...
16	25	25	25	16	...
17	20	20	20	17	...
18	15	15	15	18	...
19	10	10	10	19	...
20	5	5	5	20	...

```

; NAME/KEY: CDS
; LOCATION: (52) ... (2259)
US-10-045-815-3
Alignment Scores:
Pred. No.: 0 Length: 2600
Score: 3870.00 Matches: 736
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-10-045-815-4 (1-736) x US-10-045-815-3 (1-2600)
Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20
Db 52 ATGCGGCTACGCGGTTGAAGCTGCTGACCACTGCTGGCTGCTGGCGGCTGCTCC 111
Qy 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40
Db 112 CAAGCCGAGGTCGAGTCCGAGGCGAGATGGGCAATGGTGAAGCTGATCTGCTTCGCC 171
Qy 41 GluGlyThrAlaAlaLysAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60
Db 172 GAGGGGACCGAGCTTACGCGCGCGGGGACTGGCCCGGGTGTCTCTGAGCATGGAAACGG 231
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
Db 232 GCGTGTGGCTCCCGGGGAGCGCTCCGCGCGCTTTCGCTGGCGCTCCGACCCAGTGTGCC 291
Qy 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100
Db 292 GCCGACTTCCGTCGAGCTGGACCCCGACTGTGTCGCCCGCGCGCCGAGGCTCGGGC 351
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLeu 120
Db 352 GCCGCGCGCTCGCGGACCTGAGCTTCTCGGGGGCTTCTGCGTGGCGCTGCTGCTG 411
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140
Db 412 CGCCGCTGCTCGGGCGCGCGCGCCGCTGCTGCTCAGCGAAGATGGAGCTGGAGTTC 471
Qy 141 ArgLysArgSerProTrpArgAsnTrpLeuGlnValAlaTrpPheLysLeuAsnLysLeuGlu 160
Db 472 CGCAAGCGGAGCCCTACAACCTACTGCTGAGTGGCTTCTCAAGATCAACAAGTTGGAG 531
Qy 161 LysAlaValAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180
Db 532 AAAGCTGTGCTGAGCAGACACCTTCTCGTGGGCAATCTGAGCAGATGGAAATGCAG 591
Qy 181 GlnAsnLeuAspTrpTrpGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200
Db 592 CAGAACCTAGACTATTACCAACCATCTCTGGAGTGAAGCGCGGACTTCAAGGATCTT 651
Qy 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTrpSerGluGluGln 220
Db 652 GAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTGGACTCTACTCAGAGGAACAG 711
Qy 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTrpPheValAlaTrpGlu 240
Db 712 CCACAGGAGCTGTGCCCCACCTAGAGCGCGGCTGCAAGAAATACITTTGGGCTATGAG 771
Qy 241 GluCysArgAlaLeuCysGluGlyProTrpAspTrpAspGlyTrpAsnTrpLeuGluTrp 260
Db 772 GAGTGGCGTGCCTCTCGAAGGCGCCCTATGACTACGATGGCTGAGTCTACTACCTTGGATC 831
Qy 261 AsnAlaAspLeuPheGlnAlaHisThrAspHisTrpIleGlnValLeuAsnCysLysGln 280
Db 832 AACGCTGACCTCTCCAGGCGCATCACAGATCATATCATCCAGTCTCTCACTGATGACGAG 891
Qy 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300
Db 892 AACTGTGTACGAGCTGTGCTCCACCCCAAGTCGAGAGAAGCCCTTTGAAGACTTCCCTC 951

```

---

```

301 ProSerHisTrpAsnTrpLeuGlnPheAlaTrpTrpAsnIleGlyAsnTrpThrGlnAla 320
952 CCATCGCATTAATTAATCTCGAGTTTGCCTACTATAACATTTGGAAATTATACAAAGCT 1011
Qy 321 GlyGluCysAlaLysThrTrpLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340
1012 GGTGAATGTGCAAGACCTATCTCTCTTCTCCCAATGACGAGGTGATGAACCAAAAT 1071
341 LeuAlaTrpTrpAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360
1072 TTGGCTTATATGAGCTATGCTTGGAGAGAACACACACCATCCATCGGCCCGCTGAG 1131
Qy 361 SerAlaLysGluTrpArgGlnArgSerLeuLeuLysGluLeuLeuPhePheAlaTrp 380
1132 AGTGCACAGGAGTACCCGACAGCAAGCTACTGTGAAAAGAACTGCTTTCTTCGCTTAT 1191
381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400
1192 GATGTTTGGATTTCCCTTTGTGATCCGATTCATGGAATCCAGAGAAGTATGATCCC 1251
401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValAlaArgIleSerGlnGlu 420
1252 AAGAGATTGCAAGAGAAACAGAAAGTCAGAACGGGAAACAGCCGTACGCATCTCCAGGAG 1311
421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440
1312 ATTGGAAACCTTTATGAGAAATCGAGACCTTTGTGGAAGAGAAAGACCAAGGAGTCACTG 1371
441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTrpGluGlyIleSerLeuThr 460
1372 GATGTGAGCAGACTGACCCGGGAGGTGGCCCTGCTGTATGAAGCAATCAGTCTCACC 1431
461 MetAsnSerLysLeuLeuAsnGlyTrpGlnArgValValMetAspGlyValIleSerAsp 480
1432 ATGAATCCAAACTCCTGAATGGTTACGAGCGGTGGTATGGACGGCGTAAATCTCTGAC 1491
481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTrp 500
1492 CACGAGTGTGAGGAGTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGCTAC 1551
501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTrpGlyValThrValPheLys 520
1552 CGGGGTGAGACTCCCCACATCTCCCAATGAAAGTTCTATGTGTCTACTGTCTTCAAA 1611
521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTrpTrpAsn 540
1612 GCCCTCAAGCTGGGGCAGAGGCAAGTTCTCTGCGAGAGTGGCCACCTGTACTACAAC 1671
541 ValThrGluLysValArgArgIleMetGluSerTrpPheArgLeuAspThrProLeuTrp 560
1672 GTGACGAGAAAGTGGCGGCTCATGAGTCTTCTTCCGCTGGATACGCCCTCTAC 1731
561 PheSerTrpSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580
1732 TTTTCTCTACTCTATCTGTGTGCGCACTGCCATCGAAGAGGTCCAGGAGAGAGAG 1791
581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600
1792 GATGATAGTATCCAGTCCAGTCCAGCAACATGTCATCTGAATGCGGAGACCTCGTGTGT 1851
601 ValLysGluProProAlaTrpThrPheArgAspTrpSerAlaIleLeuTrpLeuAsnGly 620
1852 GTCAAGAGCCCCCAGCCCTACACCTTCCGCGACTACAGCGCATCTTTACCTTAATGGG 1911
621 AspPheAspGlyGlyAsnPheTrpThrGluLeuAspAlaLysThrValThrAlaGlu 640
1912 GACTTCGATGCGGAAACTTTTATTTTCACTGAATGATGCCAAGACCGTGACGACAGAG 1971
641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660
1972 GTGCGCCCTCAGTGTGGAAGAGCGGTGGATTTCTTTCAGGCACTGAAACCCACATGGA 2031

```



QY 661 ValIysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680  
Db 2032 GTGAAGGCTGTACACAGGGGCGAGCGGTGTGCCATCGCCCTGTGGTTACCCCTGGACCT 2091  
QY 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValIysMetLeuPheSerPro 700  
Db 2092 CGACACAGCGAGCGGACAGGGTGCAGGAGATGACCTGGTGAAGATGCTTTTCAGCCCA 2151  
QY 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProGluPro 720  
Db 2152 GAAGAGATGACCTCTCCAGGAGCAGCCCTGGATGCCAGCAGGGCCCCCGAACCT 2211  
QY 721 AlaGlnGluSerLeuSerGlySerGluSerIysProIysAspGluLeu 736  
Db 2212 GCACAGAGTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2259

RESULT 2  
US-10-302-172-228  
; Sequence 228, Application US/10302172  
; Publication No. US20040053250A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. US20040053250A1el Arginine-rich Protein-like Nucleic Acids an  
; FILE REFERENCE: Polypeptides  
; CURRENT APPLICATION NUMBER: US/10/302,172  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR FILING DATE: 2002-08-20  
; PRIOR APPLICATION NUMBER: PCT US02/05095  
; PRIOR FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: US 09/799,451  
; NUMBER OF SEQ ID NOS: 950  
; SOFTWARE: pt FL\_genes Version 2.0  
; SEQ ID NO 228  
; LENGTH: 2242  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (32)..(2239)  
US-10-302-172-228

Alignment Scores:  
Pred. No.: 0 Length: 2242  
Score: 3861.00 Matches: 734  
Percent Similarity: 99.86% Conservative: 1  
Best Local Similarity: 99.73% Mismatches: 1  
Query Match: 99.77% Indels: 0  
DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x US-10-302-172-228 (1-2242)

QY 1 MetAlaValArgAlaLeuIysLeuLeuThrThrLeuAlaValAlaAlaLaser 20  
Db 32 ATGGCGGTACCGCGTTGAGCTGTGACCACTGCTGGTGTGCGCCCTGGCTCC 91  
QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40  
Db 92 CAAGCCGAGTCCAGTCCGAGGAGGATGGGCGATGGTGCAGCCCTGATCTGCTCTGGCC 151  
QY 41 GluGlyThrAlaAlaValArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
Db 152 GAGGGACCGCAGCCCTACTTCGCGGGGAGCTGGCCGGGGTGGTCTCGAGCATGGAACGG 211  
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
Db 212 GCGCTCGGCTCCGCGGAGCCCTCCGCGCCCTTCGCTCGGCTGGCGCACCAGGTGGCC 271  
QY 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProAlaGlnAlaSerGly 100

Db 272 GCCGACTTCCGTTGGAGCTGGACCCGACTGGTCCCGCCAGCCGCGCCAGGCTTCGGGC 331  
QY 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
Db 332 GCCCGCGCCCTGCGGAGCTTCTTTCGGGGGCGCTTTCGGCGTGGCTGCGCTGCGCTG 391  
QY 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
Db 392 CGCGCTGCTCGGCGCGCGCCGCTCGCTCAGCGAAGAGATGAGGTGGAGTTC 451  
QY 141 ArgIysArgSerProTyrAsnTyrLeuGlnValAlaValPheLysIleAsnLysLeuGlu 160  
Db 452 CGCAAGCGGAGCCCTACAACTACCTGCGAGGTGCGCTACTTCAAGATCAACAAGTTGGAG 511  
QY 161 LysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetGln 180  
Db 512 AAAGCTGTGTGTCGAGCACACACTTCTTCTGGGCAATCTCTGAGCACATGAAATGCG 571  
QY 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValIysGluAlaAspPheLysAspLeu 200  
Db 572 CAGAACCTAGACTATTACCAAAACATGTCTGGAGTGAAGGAGGCGGACTTCAAGGATCTT 631  
QY 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220  
Db 632 GAGACTCAACCCCATATCAAGAATTTGACTGGGAGTGGGACTCTACTCAGAGGAACAG 691  
QY 221 ProGlnGluAlaValProHisGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
Db 692 CCACAGGAGGTGTGCCCCACCTAGAGCGCGGCTGCAAGATACTTTGTGGCTATGAG 751  
QY 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
Db 752 GAGTCCGCTGGCTCTGCGAAGGCGCCCTATGACTACGATGCGTACAACTACCTTGGATAC 811  
QY 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysIysGln 280  
Db 812 AACGCTGACCTCTTCCAGGCGCATCACAGATCATTCACAGGTCCTCATCTGTAAAGAG 871  
QY 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
Db 872 AACTGTGTACGGAGCTTGTCTCCACCAAGTCAGAGAAGCGCTTTTGAAGACTTCTCTC 931  
QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTrpAsnIleGlyAsnTyrThrGlnAla 320  
Db 932 CCATCGCATTAATATTCTGCGAGTTTGCCTACTATACATTGGGAATTTATACAAAGCT 991  
QY 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
Db 992 GGTGAATGTGCAAGACCTACTTCTCTTCTTCCCAATGACGAGGTGATGAACCAAAAT 1051  
QY 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
Db 1052 TTGGCCTATTATGCGAGCTATGCTTTGGAGAAGAACACACAGATCCATCGGCCCCCGTGAG 1111  
QY 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr 380  
Db 1112 AGTGCAAGGAGTACCGACGAGGAGCTACTGGAAGAAGAACTGCTTTTCTTGGCTTAT 1171  
QY 381 AspValPheGlyTyrLeuProPheValAspProAspSerTrpThrProGluGluValIlePro 400  
Db 1172 CATGTTTTTGGAAATTCCTTTTGTGATCCGATTCATGGACTCCAGAGAGAGTGAATCCC 1231  
QY 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
Db 1232 AAGAGATTGCAAGAGAACAGAGTCAGAGCGGAACAGCGGATCCGATCTCCAGAGAG 1291  
QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
Db 1292 ATTGGGAACCTTATGAAGGAATTCAGAGCCCTTGTGGAAGAGAAGACCAAGGAGTCACTG 1351  
QY 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460

```
Db 1352 GATGTGACAGACTACCGGGAAGTGGCCCTCTGTATGAAGCATCAGTCTCACC 1411
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValMetAspGlyValIleSerAsp 480
Db 1412 ATGACTCCAACTCCTGATGTTACCAAGGGTGTGATGACGGCGGTATCTCTGAC 1471
Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500
Db 1472 CACGAGTGTGAGGAGTGTGACAGACTGACCAATGTGGCAGCAACCTCAGGAGATGCTAC 1531
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520
Db 1532 CGGGGTGAGACCTCCCACTATCTCCCAATGAAGAAGTTCATGTGTCTCACTGTCTCAA 1591
Qy 521 AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540
Db 1592 GCCCTCAAGCTGGGCAAGAAGCAAGTTCCTCTGAGAGTGCACACCTGTACTACAAC 1651
Qy 541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560
Db 1652 GTGACGGAGAAAGTGGCGGCATCATGGAGTCTCTTCCGCTGGATACGCCCTCTCTAC 1711
Qy 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArgLys 580
Db 1712 TTTTCTCTACTCTATCTGGTGTGGCGCACTGCCATCGACAGGTCCAGGCGAGAGGAAG 1771
Qy 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600
Db 1772 GATGATAGTATCCAGTCCACGTGGACAACTGATCTCTGAATCCGAGACCTCTGTGT 1831
Qy 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620
Db 1832 GTCAAGAGACCCGAGCTACACTTCCGCGACTACAGCGCCATCTTTACCTAATGGG 1891
Qy 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640
Db 1892 GACTTCGATGGCGAAACTTTATTTTCACTGAATGGATGCCAAGACCGGTGACGGCAGAG 1951
Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660
Db 1952 GTGCAGCTCAGTGTGAAGAGCGGTGGGATCTCTTCAGGCACCTGAAAACCCCAATGGA 2011
Qy 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuAspPro 680
Db 2012 GTGAAGGCTGTCCACAGGGGCGCGCTGTGCCATCGCCCTGTGGTTCACCTTGGACCT 2071
Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700
Db 2072 CGACACAGCAGCGGAGCAGGGTGCAGGAGATGACCTGTGTGAAGATGCTCTTCAGCCCA 2131
Qy 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro 720
Db 2132 GAAGAGATGACCTCTCCAGGAGCAGCCCTGGATGTCGCCCGAGGGGCCCTCCGACCT 2191
Qy 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736
Db 2192 GCACAAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGTA 2239
```

## RESULT 3

```
US-10-257-174-2
; Sequence 2, Application US/10257174
; Publication No. US20040034194A1
; GENERAL INFORMATION:
; APPLICANT: Agarwal, Pankaj
; APPLICANT: Murdock, Paul R.
; APPLICANT: Rizvi, Safia K.
; APPLICANT: Smith, Randall F.
; APPLICANT: Xiang, Zhaoying
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50022
; CURRENT APPLICATION NUMBER: US/10/257,174
; CURRENT FILING DATE: 2002-10-10
; PRIOR APPLICATION NUMBER: PCT/US01/11797
```

```
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/196,603
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/199,417
; PRIOR FILING DATE: 2000-04-24
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PstSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2211
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-257-174-2
```

```
Alignment Scores:
Pred. No.: 0 Length: 2211
Score: 3846.00 Matches: 733
Percent Similarity: 99.59% Conservative: 0
Best Local Similarity: 99.59% Mismatches: 3
Query Match: 99.38% Indels: 0
DB: 13 Gaps: 0
```

US-10-045-815-4 (1-736) x US-10-257-174-2 (1-2211)

```
Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20
Db 1 ATGGCGGTACGCGGTGAAGCTCTGACCACTGTGGCTGTCTGGCGCGCTGCTCC 60
Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla 40
Db 61 CAAGCCGAGGTGCGAGTCGAGCGAGGATGGGGCATGGTGACCGCTGATCTCTTCGCC 120
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60
Db 121 GAGGGGACCGCAGCTACGCGCGGGGACTGGCCCGGGTGGTCTCTGAGCATGGAACGG 180
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
Db 181 GCGCTGCGCTCCCGGCGAGCCCTCCGCGCCCTTCGCTCGCTCCGCGCCAGCCAGTGTGCC 240
Qy 81 AlaAspPheProTyrGluLeuAspProAspTyrSerProSerProAlaGlnAlaSerGly 100
Db 241 GCCGACTTCCCGTGGAGCTGGACCCCGACTGTGTCGCCAGCCCGCCAGGCTCGGGC 300
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120
Db 301 GCGCGCGCCCTGCGCAGCTTCTTCGCGGGGCTTCTGCGCTCGCTGCGCTGCTGCTG 360
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140
Db 361 GCGCGCTCCCTCGGGCCCGCGCCCGCCACTCGCTCAGCGAAGAGATGGAGCTGGAGTTC 420
Qy 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160
Db 421 CGAAGCGGAGCCCTTACCACTTACCTGAGGTTCGCTGCTTCAAGATCAACAAGATTGGAG 480
Qy 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180
Db 481 AAGCTGTGTGCGACACACACCTTCTTCGTGGGCAATCTCTGAGCAATCGGAATATGAG 540
Qy 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200
Db 541 CAGAACCTAGACTATTACCAACCACTTGTCTGGAGTGAAGGAGGCCGACTTCAAGGATCTT 600
Qy 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220
Db 601 GAGACTCAACCCCATATGCAAGAAATTTTCAGCTGGAGTGGGACTCTACTCAGAGGAACAG 660
Qy 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240
Db 661 CCACAGGAGGTGTGCCCCCCTAGAGGGCGCTGCAGAGATATCTTTGTGGCTATGAG 720
Qy 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGlyTyr 260
```

Db	721	GAGTGGCGTGCCCTCTCCGAAGGCCCTATGACTACGATGGCTACAACTACCTTGAGTAC	780
Qy	261	AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln	280
Db	781	AACGCTGACCTCTTCCAGGCCATCACAGATCATTTACATCCAGGTCTCTCAACTGTAAAGCAG	840
Qy	281	AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAaspPheLeu	300
Db	841	AACGTGTCTCAGGAGCTTGCTTCCACCCAAAGTCGAGAAGGCCCTTTGAAGACTTCTCCTC	900
Qy	301	ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla	320
Db	901	CCATCGCATATATATTATCTCGAGTTTGCTACTATATAACTTGGGAATTATACACAGGCT	960
Qy	321	GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn	340
Db	961	GTTGAATGTGCAAGACCTATCTTCTCTCTTCCCAATGACGAGGTGATGAACAAAAT	1020
Qy	341	LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu	360
Db	1021	TTGGCCTATTATGCAGCTATGCTTTGGAGAAGACACACCAAGTCCATCGGCCCCCGTGAG	1080
Qy	361	SerAlaLysGluTyrArgGlnArgSerLeuGluLysGluLeuLeuPhePheAlaTyr	380
Db	1081	AGTGCCAAGGAGTACCGACAGCGAAGCGCTACTGGAATAAGAACTGCTTTTCTTCGCTTAT	1140
Qy	381	AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro	400
Db	1141	GATGTTTTTGGAAATCTCCCTTTGTGGATCCGGAATTCATGGACTCCAGAAGAAGTGAITCCC	1200
Qy	401	LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu	420
Db	1201	AAGAGATTGCAAGAGAAACAGAGTCAAGACGGGAAACACGCGTACGCATCTCCAGGAG	1260
Qy	421	IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu	440
Db	1261	ATTGCGAAACCTATTGAAGGAAATCGACACCCCTGTGTGAAGAGAGAACCAAGGAGTCACTG	1320
Qy	441	AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr	460
Db	1321	GATGTGACGACACTGACCCGGAAAGTGGCCCCCTGCTGTATGAGGCATCAGTCTCACC	1380
Qy	461	MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp	480
Db	1381	ATGAACCTCCAAACTCTCTGAATGGTTCCAGCGGGTGTGTGAGCGCGCTAACTCTGTAC	1440
Qy	481	HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr	500
Db	1441	CACGAGTGTACGAGAGCTGCAGAGACTGACCAATGTGCAGCAACCTCAGAGATGGCTAC	1500
Qy	501	ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys	520
Db	1501	CGGGGTGACACCTCCCCACATACTCCCAATGAAAGTTCTATGGTGTCACTGTCTTCAAA	1560
Qy	521	AlaLeuLysLeuGlyGlnGluGlyValProLeuGlnSerAlaHisLeuTyrTyrAsn	540
Db	1561	GCCCTCAAGCTGGGGCAAGAGGCAAGTTCTCTCTGAGAGTGCCCACTGTACTACAC	1620
Qy	541	ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr	560
Db	1621	GTGACGGAGAGGTGGCGGCATCATGGAGTCTTACTTCCGCTGGATACGCCCTCTAC	1680
Qy	561	PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys	580
Db	1681	TTTTTCTACTCTCATCTGGTGTGCCCATGTGCCATCGAAGAGGTCCAGGACAGAGAAG	1740
Qy	581	AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys	600
Db	1741	GATGATGACTCATCCAGTCCACGTGGACACTGCATCTCTGAATGCCGAGACCTCTGTGT	1800
Qy	601	ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly	620
Db	1801	GTCAAGAGAGCCCCAGAGCTACACTCTCGGCACTACAGCGCACTCTTACTCTAAATGGG	1860

## RESULT 4

```

US001.1
; Sequence 69, Application US/10312352
; Publication NO. US20040053824A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.; TANG, Y. Tom
; APPLICANT: YUE, Henry; AZIMZAI, Yalda
; APPLICANT: HE, Ann; BATPA, Saajeav
; APPLICANT: LO, Terence P.; NGUYEN, Danniell B.
; APPLICANT: BURRILL, John D.; MARCUS, Gregory A.
; APPLICANT: ZINGLER, Kurt A.; GANDHI, Ameena R.
; APPLICANT: LAL, Preeti G.; KEARNEY, Liam
; APPLICANT: BURFORD, Neil; YAO, Monique G.
; APPLICANT: CHAMLA, Narinder K.; ELLIOT, Vicki S.
; APPLICANT: ARVIZU, Chandra S.; KHAN, Farrah A.
; APPLICANT: BAUGHN, Mariah R.; HAFALIA, April, J.A.
; APPLICANT: POLICKY, Jennifer L.; AU-YOUNG, Janice K.
; APPLICANT: LU, Yan; BOROWSKY, Mark L.
; APPLICANT: LU, Dyung Aina M.; RAMKUMAR, Jayalaxmi
; APPLICANT: YANG, Junming; GURURAJAN, Rajagopal
; APPLICANT: WARREN, Bridget A.; GIETZEN, Kimberly J.
; APPLICANT: XU, Yuming; KALLICK, Deborah A.
; APPLICANT: LEE, Ernestine A.; THANGAVELU, Kavitha
; APPLICANT: DELGEGANE, Angelo M.; LEE, Sally
; TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES
; FILE REFERENCE: PF-0794 USN
; CURRENT APPLICATION NUMBER: US/10/312,352
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: PCT/US01/21067
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215,454
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 60/219,462
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: US 60/240,111
; PRIOR FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US 60/240,106
; PRIOR FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: US 60/244,021
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US 60/248,887
; PRIOR FILING DATE: 2000-11-14
; PRIOR APPLICATION NUMBER: US 60/249,570
; PRIOR FILING DATE: 2000-11-16
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PERL Program
; SEQ ID NO 69
; LENGTH: 2583

```



Db 1962 GTGAGCCTCAGTGTGAAGAGCGTGGATTCTCTTCAGGCACTGAAACCCACATGA 2021  
Qy 661 VallyAlaValThrArgGlyGlnArgCysAlaAlaLeuTrpPheThrLeuAspPro 680  
Db 2022 GTGAAGGCTGTCAACAGGGGACAGCGCTGTGCCATCGCCCTGTGGTCACTCCCTGGACCT 2081  
Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMetLeuPheSerPro 700  
Db 2082 CGACACAGCGAGCGGACAGCGGTGCGAGCAGATGACTGTGTGAGATGCTCTTCAGCCCA 2141  
Qy 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProGluPro 720  
Db 2142 GAAGAGATGACCTCTCCAGGAGCAGCGCCCTGGATGCCAGCGGCCCCCGCAACCT 2201  
Qy 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2202 GCACAAAGAGTCTCTCAGGCACTGAATCGAAGCCCAAGATGAGCTA 2249

RESULT 5  
US-09-728-952-28  
; Sequence 28, Application US/09728952  
; Patent No. US20020111302A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Zhou, Ping  
; APPLICANT: Goodrich, Ryle  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Yamazaki, Vicki  
; APPLICANT: Ujwal, Manusha L.  
; APPLICANT: Dmanac, Radoje T.  
; TITLE OF INVENTION: No. US20020111302A1el Nucleic Acids and  
; FILE OF INVENTION: Polypeptides  
; CURRENT APPLICATION NUMBER: US/09/728,952  
; CURRENT FILING DATE: 2000-11-30  
; NUMBER OF SEQ ID NOS: 101  
; SOFTWARE: pt\_FL\_genes Version 2.0  
; SEQ ID NO 28  
; LENGTH: 2753  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2418)  
US-09-728-952-28

Alignment Scores:  
Pred. No.: 0 Length: 2753  
Score: 3801.00 Matches: 733  
Percent Similarity: 90.94% Conservative: 0  
Best Local Similarity: 90.94% Mismatches: 3  
Query Match: 98.22% Indels: 70  
DB: 9 Gaps: 1

US-10-045-815-4 (1-736) x US-09-728-952-28 (1-2753)

Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
Db 1 ATGGCGGTACCGCGTGTGAAGTCTGCTGACACACTGCTGGCTGTCTGGCGCTGCTCC 60  
Qy 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40  
Db 61 CAAGCCAGGTCGAGTCCGAGGAGGATGGGCGATGGTGGCTGATCTGTCTTCGCGC 120  
Qy 41 GluGlyThrAlaAlaValAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
Db 121 GAGGAGCCCGACCTACGCGCGCGGAGCTGGCCCGGGGTGTCTCTGAGCATGGACGG 180  
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80

Db 181 GCGCTGCGCTCCGGGAGAGCCCTCCGCGCCCTTCGCTGCGCTCCGCGACCCAGTGTGCC 240  
Qy 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
Db 241 GCGCACTTCCGCTGGAGCTGAGACCCAGCTGGTCCCGCCAGCCGCGCCAGCCCTCGGGC 300  
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
Db 301 GCGCGCGCCCTCGCGGACCTGAGCTTCTTCGGGGCCCTTCGCTGCGCGCTCGCTGCGCT 360  
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
Db 361 GCGCGCTGCTCCGCGCGCCGCGCCAGCTCGCTCAGCGAAGAGATGGAGCTGGAGTTC 420  
Qy 141 ArgLysArgSerProTrpAsnTyrLeuGlnValAlaAlaTyrPhe 154  
Db 421 CGCAAGCGGAGCCCTCACTACTGAGTGCCTACTTCAAGGTGCAGACCTGCGCTG 480  
Qy 154 154  
Db 481 GAACCGAGCGCGCGGGTCTTCTTGGGAGAGAGTGTTCGAGGGGACCTGAGGAGCTTG 540  
Qy 154 154  
Db 541 GGGGATCGGGGAAGTGTCCGAGGAGGGGAAAGTGGCTCTCTGCTGGGAGCTCTCT 600  
Qy 154 154  
Db 601 CGGAGCGGGAGAGAGCTCTCCCTGGCAGGAGAGACTTCTTCCGCCAGTTCGATGGGCG 660  
Qy 155 -----LysIleAsnLysLeuGluLysAlaValAlaAlaAlaHisThrPhePhe 170  
Db 661 ATGCTAACCCCAAGATCAACAAGTTGGAGAAGCTGTGTGTCGAGCAGCACACCTTCTTC 720  
Qy 171 ValGlyAsnProGluHisMetGluMetGlnGlnAsnLeuAspTyrTyrGlnThrMetSer 190  
Db 721 GTGGCAATCTCTGAGCATGGAATGCAGCAGAACCTAGACTATTACCAACCATGTCT 780  
Qy 191 GlyValLysGluAlaAspPheLysAspLeuGluThrGlnProHisMetGlnGluPheArg 210  
Db 781 CGAGTGAAGAGGCGGACTTCAAGGATCTTCAGACTCAACCCCATATGCAAGAATTTCGA 840  
Qy 211 LeuGlyValArgLeuTyrSerGluGluGlnProGlnGluAlaValProHisLeuGluAla 230  
Db 841 CTGGGAGTGGCACTCTACTCAGAGAAACAGCCAGGAAAGCTGTGCCCGCCAGCTAGAGCG 900  
Qy 231 AlaLeuGlnGluTyrPheValAlaTyrGluGluCysArgAlaLeuCysGluGlyProTyr 250  
Db 901 GCGCTGCAAGAAATACTTTGTGGCTATGAGGAGTGGCGCTCTCGGAGAGGCGCCCTAT 960  
Qy 251 AspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAlaIleThrAsp 270  
Db 961 GACTACGATGGCTACACTACTCTTGTAGTACAACTGACCTGACCTTCCAGGCCATCACAGAT 1020  
Qy 271 HisTyrIleGlnValLeuAsnCysLysGlnAsnCysValThrGluLeuAlaSerHisPro 290  
Db 1021 CATTACATCCAGGTCTCACTGTAAAGCAAGAACTGTGTACGGAGCTTGTCTCCACCCA 1080  
Qy 291 SerArgGluLysProPheGluAspPheLeuProSerHisTyrAsnTyrLeuGlnPheAla 310  
Db 1081 AGTCGAGAGAAAGCCCTTTGAAGACTTCTCCCATCGCATTAATAATTATCTCAGTTTCC 1140  
Qy 311 TyrTyrAsnIleGlyAsnTyrThrGlnAlaGlyGluCysAlaLysThrTyrLeuLeuPhe 330  
Db 1141 TACTATAACATTTGGGAATTATACACAGGCTGTGAATGTGCCAAGACCTATCTCTCTTC 1200  
Qy 331 PheProAsnAspGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGlu 350  
Db 1201 TTCCCAATGACGAGGTGATGAACCAAAATTTGGCTATTATGAGCTATGCTTTGGAGAA 1260  
Qy 351 GluHisThrArgSerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeu 370

Db 1261 GAACACACAGATCCATCCGCCCCCGTGGAGTGCCCAAGGAGTACCACAGCGAAGCCTA 1320  
Qy 371 LeuGluLysGluLeuLeuPheAlaTyrAspValPheGlyIleProPheValAspPro 390  
Db 1321 CTGGAAGAAAGAACTGCTTTCTTCGCTATGATGTTTTCGAATTCCTTTGGATCCG 1380  
Qy 391 AspSerTrpThrProGluGluValIleProLysArgLeuGlnGluLysGlnLysSerGlu 410  
Db 1381 GATTCATGGACTCCAGAGAGAGTATCCCAAGAGATTGCAAGAGAAACAGAGAGTCAGAA 1440  
Qy 411 ArgGluThrAlaValArgIleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThr 430  
Db 1441 CGGAAACAGCGCTGACGATCTCCAGAGAGATTGGGAACCTTATGAAGGAAATCGAGACC 1500  
Qy 431 LeuValGluGluLysThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyGly 450  
Db 1501 CTTGTGAAGAGAGACCAAGAGTCTACTGATGTGAGCAGACTGACCCCGGAGAGTGGC 1560  
Qy 451 ProLeuLeuTyrGluGlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyrGln 470  
Db 1561 CCCCTGCTGATGAAGCATCAGTCTCACCATGAATCCAACTCTCTGAATGCTTCCAG 1620  
Qy 471 ArgValValMetAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThr 490  
Db 1621 CGGTGTGTGATGGACGGCGTAATCTCGACCAAGAGTGTGAGGAGCTGCAGAGCTGAC 1680  
Qy 491 AsnValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsn 510  
Db 1681 AATGTGCAGCAACCTCAGAGAGTGTCTACCGGGTCCAGACCTCCCCACATCTCCCAAT 1740  
Qy 511 GluLysPheTyrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGlyLysVal 530  
Db 1741 GAAAGAGTCTATGTGTGCTACTGCTTCAAAGCCCTCAAGCTGGGGCAAGGCAAGATT 1800  
Qy 531 ProLeuGlnSerAlaHisLeuTyrTyrAsnValThrGluLysValArgAlaGlyMetGlu 550  
Db 1801 CCTCTGAGAGTGCACCTGTACTACACGTGACGGAGAGGTGGCGGCATCATGGAG 1860  
Qy 551 SerTyrPheArgLeuAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThr 570  
Db 1861 TCCTACTTCGCCCTGGATAGCCCTCTACTTTCTCTACTCTCATCTGCTGCTGCCGACCT 1920  
Qy 571 AlaIleGluGluValGlnAlaGluArgLysAspSerHisProValHisValAspAsn 590  
Db 1921 GCCATCGAAGAGGTCCAGGACGAGAGAGATGATGATCATCATGCTCCACGTGGACAC 1980  
Qy 591 CysIleLeuAsnAlaGluThrLeuValCysValLysGluProProAlaTyrThrPheArg 610  
Db 1981 TGCATCTGTAATGCCAGAGACCTCGTGTGTCAAAGAGCCCCAGCCTACACCTTCCGC 2040  
Qy 611 AspTyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThr 630  
Db 2041 GACTACAGCGCCATCTCTTAATATGGGAGCTTCGATCGCGGAACTTTTATTTCACCT 2100  
Qy 631 GluLeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGly 650  
Db 2101 GAACCTGGATGCCAAGACCGTGACGGCAGAGGTGCAGCTTCAGTGTGAAGAGCCGTGG 2160  
Qy 651 PheSerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCys 670  
Db 2161 TTCTCTTCAGGACATGAAACCCACATGAGTGAAGCTCTGTACCAAGGGGGAGCGCTGT 2220  
Qy 671 AlaIleAlaLeuTyrPheThrLeuAspProArgHisSerGluArgAspArgValGlnAla 690  
Db 2221 GCCATCGCCCTGTGTTTACCTTGACCTCTGCACAGCGAGCGGACAGGGTGCAGGCA 2280  
Qy 691 AspAspLeuValLysMetLeuPheSerProGluGluMetAspLeuSerGlnGluGlnPro 710  
Db 2281 GATGACCTGGTGAAGATGCTCTTCAGGCCCAAGAGAGATGAGACCTCTCTCCAGGAGAG 2340  
Qy 711 LeuAspAlaGlnGlnGlyProProGluProAlaGlnGluSerLeuSerGlySerGluSer 730  
Db 2341 CTGGATGCCCAAGAGGGGCCCCCGAACCCTGCAACAGATCTCTCTCAGGCACTGATGTC 2400

Qy 731 LysProLysAspGluLeu 736  
Db 2401 AAGCCCAAGGATGAGCTA 2418

## RESULT 6

US-09-728-952-51  
; Sequence 51, Application US/09728952  
; Patent No. US20020111302A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Zhou, Ping  
; APPLICANT: Goodrich, Ryle  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Yamazaki, Vicki  
; APPLICANT: Ujwal, Manusha L.  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. US20020111302A1el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 799  
; CURRENT APPLICATION NUMBER: US/09/728,952  
; CURRENT FILING DATE: 2000-11-30  
; NUMBER OF SEQ ID NOS: 101  
; SOFTWARE: pt\_FL\_genes Version 2.0  
; SEQ ID NO 51  
; LENGTH: 2753  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2418)  
US-09-728-952-51

Alignment Scores:  
Pred. No.: 0 Length: 2753  
Score: 3801.00 Matches: 733  
Percent Similarity: 90.94% Conservative: 0  
Best Local Similarity: 90.94% Mismatches: 3  
Query Match: 98.22% Indels: 70  
DB: 9 Gaps: 1

US-10-045-815-4 (1-736) x US-09-728-952-51 (1-2753)

Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
Db 1 ATGGCGGTACGCGGCTTGAAGCTGTGACACACACTGCTGGCTGTCTGGCGCGCTGCCTCC 60  
Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuPheAla 40  
Db 61 CAAGCCGAGGTCGAGTCGCGAGCAGGATGGGGCATGGAGCGCTGATCTCTCTTCCGCC 120  
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
Db 121 GAGGGGACCGCAGCGCTACGCGCGGGGACCTGGCGGGTGGTCTTGAGCATGGAACGG 180  
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
Db 181 GCGCTGCGCTCCCGGGCAGCCCTCCGCGCCCTCTGCGCTGCGCTCCGCGCACCCAGTGTGCC 240  
Qy 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
Db 241 GCCGACTTCCCTGGAGCTGACCCCGGACCTGGTCCCGCCCGCCGCGCCAGGCTCGGGC 300  
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
Db 301 GCGCGCGCCCTTCGCGACCTGAGCTTCTTCGGGGGCGCTTCTGCGCTGCGCTGCGCTGCTG 360  
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
Db 361 GCGCGCTGCTTCGGGGCGCGCGCGCCCGCCACTCGCTCAGCGAAGAGATGGAGCTGGAGTTC 420

Qy	141	ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPhe	154
Db	421	CGCAAGCGGAGCCCTACAACTACCTGCAGGTGCCTTCTCAAGTGCAGAGCTGCCTG	480
Qy	154	-----	154
Db	481	GAACCCAGGCGCGGGGTCTCTCTGGGAGAGAGTGTTCACGGGACCTGAGGAGCTTG	540
Qy	154	-----	154
Db	541	GGGATCGGGGAAGTGTTCGACGGAGGGGAAAGTGGCTCTCTGGCTGGGGAGCTCTCCT	600
Qy	154	-----	154
Db	601	CGGAGCGGGAGAGAGTGCTCCCTGGGAGGAGACCTTCTCGCCAGATTTCGATGGGCGAG	660
Qy	155	-----LysIleAsnLysLeuGluLysAlaValAlaAlaHisThrPhePhe	170
Db	661	ATGCTAACCCCAAGATCAAAAGTTGGAGAAAGTGTCTGTCAGCACACACTTCTTTC	720
Qy	171	ValGlyAsnProGluHisMetGluMetGlnGlnAsnLeuAspTyrTyrGlnThrMetSer	190
Db	721	GTGGCAATCCTGAGCACATGGAAATGCACGAACTAGACTATTACCAAAACATGTCT	780
Qy	191	GlyValLysGluAlaAspPheLysAspLeuGluThrGlnProHisMetGlnGluPheArg	210
Db	781	GGAGTGAAGGAGCGCGACTTCAAGGATCTTGAGACTCAACCCCATATGCAAGAATTTCGA	840
Qy	211	LeuGlyValArgLeuTyrSerGluGluGlnProGlnGluAlaValProHisLeuGluAla	230
Db	841	CTGGGAGTGGCACTTACTCAGAGGAACACGCCACAGGAAGCTGTGCCCACTTAGAGCGG	900
Qy	231	AlaLeuGlnGluTyrPheValAlaTyrGluGluCysArgAlaLeuCysGluGlyProTyr	250
Db	901	CGCTGCACAAATACTTTGTGGCTATGAGGAGTGGCTGCCCTCTGCGAAGGCGCCTAT	960
Qy	251	AspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAlaIleThrAsp	270
Db	961	GACTACATGGCTACAACTACCTTGAGTACAAACGCTGACCTCTTCCAGGCCATCAGAT	1020
Qy	271	HisTyrIleGlnValLeuAsnCysLysGlnAsnCysValThrGluLeuAlaSerHisPro	290
Db	1021	CATTACATCCAGGTCTCACTGTAAGCAAACTGTGTCTACGGAGCTTGCTTCCCACCA	1080
Qy	291	SerArgGluLysProPheGluAspPheLeuProSerHisTyrAsnTyrLeuGlnPheAla	310
Db	1081	AGTCGAGAGAAGCCCTTGAAGACTCTCTCCCATCGCATTTATAATTATCTCGAGTTGCC	1140
Qy	311	TyrTyrAsnIleGlyAsnTyrThrGlnAlaGlyGluCysAlaLysThrTyrLeuLeuPhe	330
Db	1141	TACTATAACATTGGGAATTATACACAGCTGTGGAATGTCCAGAGCTATCTTCTTTC	1200
Qy	331	PheProAsnAspGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGlu	350
Db	1201	TTCCCCAATGACGAGGTGATGAACCAAAATTTGGCCCTATTATGCAGCTATGCTTGAGAA	1260
Qy	351	GluHisThrArgSerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeu	370
Db	1261	GAACACACCAGATCCATCGGCCCCCGTGGAGTGCACAGAGTAGTACCAGACAGCAGCTTA	1320
Qy	371	LeuGluLysGluLeuLeuPhePheAlaTyrAspValPheGlyIleProPheValAspPro	390
Db	1321	CTGGAAAAAAGAACTGCCTTTTCTCGCTTATCATGTTTGTGAATTCCTCTTGTGGATCCG	1380
Qy	391	AspSerThrProProGluGluValIleProLysArgLeuGlnGluLysGlnLysSerGlu	410
Db	1381	GATTTCATGGACTCCAGAGAAGTAGTATCCCAAGAGATTGCAGAGAAACAGAGAGTCAGAA	1440
Qy	411	ArgGluThrAlaValArgIleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThr	430
Db	1441	CGGGAACAGCCGTACCACTCTCCAGAGATTGGAACTTATGAAGAAATCGAGACC	1500

RESIST. 7

```

US000045-815-1
; Section 1 Application US/10045815
; Publication No US20020160438A1
; GENERAL INFORMATION:
; APPLICANT: Wadhwa, Renu
; APPLICANT: Sugihara, Takashi
; APPLICANT: Ohide, Akiko
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE

```







Db 1792 CTCCACATACCTCCCAATGAAGAAGTTCTATGGTGTCACTGTCTTCAAAGCCCTCAAGCT 1851  
Qy 524 uGlyGlnGluGlyValProLeuGlnSerAlaHisLeuTyrTyrAsnValThrGluGly 544  
Db 1852 GGGCAAGAAGGCAAAAGTTCTCTGAGAGTGCACCTGTACTCAACAGTCAAGAGAA 1911  
Qy 544 sValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyrPheSerTyrSe 564  
Db 1912 AGTGGCGGCATCATGGAGTCTTCTCCCTGGATAGCCCTCTCTCTTCTCTACTC 1971  
Qy 564 rHisLeuValCysArgThrAlaIleGluGlnValGlnAlaGluArgLysAspSerHis 584  
Db 1972 TCATCTGGTGTGGCGCACTGCCATCGAAGAGGTCCAGGAGAGAGAGATGATGCA 2031  
Qy 584 sProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCysValLysGluPr 604  
Db 2032 TCCAGTCCACGTGCAGCAACTGCATCTGAATGCCGAGACCTCGTGTGTGTCAAGAGCC 2091  
Qy 604 oProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGl 624  
Db 2092 CCAGAGCTACACCTTCGGGACCTACAGGCCATCCTTTACCTAAATGGGACTTCGATGG 2151  
Qy 624 yGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGluValGlnProGl 644  
Db 2152 CGAAACCTTTTATTTCACTGAATGGATGCCAAGACCGTGACGGCAGAGGTGCAGCCTCA 2211  
Qy 644 nCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGlyValLysAlaVa 664  
Db 2212 GTGTGAAGAGCCGCTGGATTCCTTCAGGCACCTGAACCCACATGGAGTGAAGCTGT 2271  
Qy 664 lThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuAspProArgHisSerGl 684  
Db 2272 CACAGGGGGCAGCGGTGCCATCGCCTGTGGTTCACCTGGACCCCTCGACACACGCA 2331  
Qy 684 uArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerProGluMetAs 704  
Db 2332 GCGGAGCAGGGTCAGCAGATGACCTGGTGAAGATGCTCTTCAGGCCCAAGAGATGGA 2391  
Qy 704 pLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluProAlaGlnGluSe 724  
Db 2392 CCTCTCCAGGAGAGCCCTGGATGCCAGCAGAGGGCCCCCGAACCTCGACAGAGTC 2451  
Qy 724 rLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2452 TCTCTCAGGCAGTGAATCGAAGCCCAAGGATGAGCTA 2488

RESULT 8  
US-10-302-172-227  
; Sequence 227, Application US/10302172  
; Publication No. US20040053250A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Dwanac, Radoje T.  
; TITLE OF INVENTION: No. US20040053250A1el Arginine-rich Protein-like Nucleic Acids an  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 803\_ICNCP  
; CURRENT APPLICATION NUMBER: US/10/302,172  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 10/225,251  
; PRIOR FILING DATE: 2002-08-20  
; PRIOR APPLICATION NUMBER: PCT US02/05095  
; PRIOR FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: US 09/799,451  
; PRIOR FILING DATE: 2001-03-05  
; NUMBER OF SEQ ID NOS: 950  
; SOFTWARE: pt FL\_genes Version 2.0  
; SEQ ID NO 227  
; LENGTH: 2152  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS

; LOCATION: (32)...(2149)  
US-10-302-172-227  
Alignment Scores:  
Pred. No.: 0  
Score: 3645.00  
Length: 2152  
Matches: 699  
Percent Similarity: 95.24%  
Conservative: 2  
Best Local Similarity: 94.97%  
Mismatch: 5  
Query Match: 94.19%  
Indels: 30  
Gaps: 13  
US-10-045-815-4 (1-736) x US-10-302-172-227 (1-2152)  
Qy .1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValValAlaAlaAser 20  
Db 32 ATGGCGTACCGCGGTGAAGCTGCTGACACACACTGCTGGCTGTGTCGGCGCTCC 91  
Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla 40  
Db 92 CAAGCCGAGGTGCGATCCGAGCGAGGATGGCGCATGTGTGACGCTGATCTCTCTGCC 151  
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60  
Db 152 GAGGGACCGCAGCTTACCGCGCGGAGACTGGCCGGGTGGTCTCTGAGCATGGAACGG 211  
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
Db 212 GCGCTGCGCTCCCGCGCAGCCCTCCGCGCCCTTCGCTGCGCTGCGCCGACAGTGTGCC 271  
Qy 81 AlaAspPheProTyrGluLeuAspProAspTyrPheSerProAlaGlnAlaSerGly 100  
Db 272 GCGGACTTCCGCGGAGCTGACCCCGACCTGGTGTCCCGCCAGCCCGCCAGGCTCGGGC 331  
Qy 101 AlaGlyAlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
Db 332 GCGCGCGCTCGCGACCTGAGCTTCTTCGGGGGCTCTTCGCTGCGCTGCGCTGCGCTG 391  
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
Db 392 GCGCGCTGCTCGCGCGCGCGCCACCTCGCTGCGAGAGAGATGAGCTGAGGATTC 451  
Qy 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
Db 452 CGAAGCGAGCGCCCTCAACTACTGCGAGTTCGCTTCAAGATCAACAGTTGGAG 511  
Qy 161 LysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetGln 180  
Db 512 AAAGCTGTGTGCAGACACACACCTTCTTCGCGCAATCTCGAGCATGGAATGCGAG 571  
Qy 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
Db 572 CAGAACCTTAGACTATTACCAACCATGTCTGAGTGAAGAGGCGGAGCTTCAAGATCTT 631  
Qy 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGln 220  
Db 632 GAGACTCAACCCCATATCAAGAAATTTTCGAGTGGAGTGGAGTCTACTCAGAGGAACAG 691  
Qy 221 ProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
Db 692 CCACAGGAAGTGTGCCCCACTAGAGCGCGCTGCAAGAAATACCTTTGTGGCTTATGAG 751  
Qy 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
Db 752 GAGTCCCGTGGCTCTGCAAGGGCCCTATGACTACGATGGCTACACTACTCTGAGTAC 811  
Qy 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
Db 812 AACGCTGACCTTCTCCAGGCCATCACAGATCATTTACATCCAGGTCTCTCACTGTAAGCAG 871  
Qy 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
Db 872 AACTGTGTCAGGAGCTTGTCTTCCACCACAGTTCGAGAGAGGCCCTTTGAGAGACTTCTCTC 931

QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
DB 932 CCATCGCAATTATAATATCTCGAGTTGGCCCTACTATAACATTGGGAATTATACACAGGCT 991  
QY 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
DB 992 GTTGAATGTCGAAGACCTATCTCTCTCTTCCCAATGACGAGGTGATGAACCAAAAT 1051  
QY 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluLysIleThrArgSerIleGlyProArgGlu 360  
DB 1052 TTGCCTATTATGAGCTATGCTTGGAGAGAACACACCATCCATCGGCCCCCGTGAG 1111  
QY 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaTyr 380  
DB 1111 ----- 1111  
QY 381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400  
DB 1112 -----GATTTCATGGACTCCAGAGGAAGTGAATCC 1141  
QY 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
DB 1142 AAGAGATTGCAAGAGAAACAGAAATCAGACGGGAACAGCCGATCCATCTCCAGAG 1201  
QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
DB 1202 ATTGGAACTTATGAAGAAATCGAGACCTTGTGAAGAGAACCAAGGAGTCACTG 1261  
QY 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
DB 1262 GATGTGACAGACTGACCCGGAAGGTGGCCCTCTGTATGAAGCATTAGTCTCACC 1321  
QY 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
DB 1322 ATGAATCCAAACTCTTGTGTTCCAGCGGTGTGTATGGCGCGTAACTCTGAC 1381  
QY 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
DB 1382 CACCAATGTACAGAGCTGCAGAGACTGACCAATGGGGCAGCAACCTCAGGAGATGGCTAC 1441  
QY 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
DB 1442 CGGGGTGAGACTCTCCACATATCTCCCAATGAAGAAGTTCTATGGTGTCACTGTTCAA 1501  
QY 521 AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
DB 1502 GCCCTCAAGCTGGGGCAAGAGGCAAAAGTCTCTGACAGAGTGCCACCTGTACTACAAC 1561  
QY 541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
DB 1562 GTGACGAGAAAGTGGCGCGCATCATGGAGTCTTCTCCCTGGATAGCCCTCTAC 1621  
QY 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArgLys 580  
DB 1622 TTTTCCTACTCTCATCTGTTGGTGGCCACTGCCATCGAAGAGTCCAGGCGAGAGGAAG 1681  
QY 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
DB 1682 GATGATAGTATCCAGTCCAGTGGCACTGATGATCTGTAATCCGAGACCCCTGTTGT 1741  
QY 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
DB 1742 GTCAAGAGCCCCAGCCTACCTCCGCGACTACAGCCCATCTTTTACCTAAATGG 1801  
QY 621 AspPheAspGlyGlyAsnPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
DB 1802 GACTTCATGCGGAAACTTTTATTTTCATCTGATGCCAAGACCGGTGACGGCAGAG 1861  
QY 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
DB 1862 GTGCAGCTCAGTGTGAAGAGCGGTGGATTCTCTTCAGGCACTGAAAACCCACATGA 1921  
QY 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuAspPro 680

DB 1922 GTGAGAGCTGTCCACAGGGGGCAGCGCTGTGCCATGCCCTGTGTTTCCCTTGGACCT 1981  
QY 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
DB 1982 CGACACAGCGAGCGGACAGGGGTGACGAGATGACCTGTGTGAAGATGCTCTTCCAGCCA 2041  
QY 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProGluPro 720  
DB 2042 GAAGAGATGAGCTCTCCAGGAGCAGCCCTGGATGCCAGAGGGCCCCCGAACCT 2101  
QY 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
DB 2102 GCACAAGAGTCTCTCTCAGGCACTGAATCGAAGCCCAAGGATGAGCTA 2149  
RESULT 9  
US-10-257-174-1  
; Sequence 1, Application US/10257174  
; Publication NO. US20040034194A1  
; GENERAL INFORMATION:  
; APPLICANT: Agarwal, Pankaj  
; APPLICANT: Murdoch, Paul R.  
; APPLICANT: Rizvi, Safia K.  
; APPLICANT: Smith, Randall F.  
; APPLICANT: Xiang, Zhaoying  
; TITLE OF INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GP50022  
; CURRENT APPLICATION NUMBER: US/10/257,174  
; CURRENT FILING DATE: 2002-10-10  
; PRIOR APPLICATION NUMBER: PCT/US01/11797  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/196,603  
; PRIOR FILING DATE: 2000-04-13  
; PRIOR APPLICATION NUMBER: 60/199,417  
; PRIOR FILING DATE: 2000-04-24  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 2127  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-257-174-1  
Alignment Scores:  
Pred. No.: 0 Length: 2127  
Score: 3572.00 Matches: 691  
Percent Similarity: 94.57% Conservative: 5  
Best Local Similarity: 93.85% Mismatches: 12  
Query Match: 92.30% Indels: 28  
DB: 13 Gaps: 3  
US-10-045-815-4 (1-736) x US-10-257-174-1 (1-2127)  
QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
DB 1 ATGGCGGTACGCGGCTTGAAGCTGTGACCACTGCTGGCTGTCTGGCGGCTGCTCC 60  
QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40  
DB 61 CAAGCCGAGGTGAGTCCGAGGAGGATGGGCACTGGGCTGTGCTGTCTTTCGCC 120  
QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
DB 121 GAGGGGACCGCAGCTACGCGCGGGGACTGGCCCGGGTGGTCTTGAGCATGGAACGG 180  
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
DB 181 GCGTGTGCGCTCCCGGAGCCCTCCCGGCCCTTCCGCTGCGCTGCGCACCCAGTGTGCC 240  
QY 81 AlaAspPheProTyrGluLeuAspProAspTrpSerProAlaGlnAlaSerGly 100  
DB 241 GCCGACTTCCCGTGGAGCTGGAACCCGACTGTGTCTCCCGAGCCCGCCGCTCGGGC 300

Qy	101	AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLeu	120
Db	301	GC CGCGCGCGCGACCTGAGCTTCTTCGGGGGCGCTTCTGCGTCGCGCTGCGCTG	360
Qy	121	ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe	140
Db	361	CGCGCTGCGCTCGGGCGCGCGCGCGCCACTCGCTCAGCGAAGAGATGAGCTGGAGTTC	420
Qy	141	ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu	160
Db	421	CGCAAGCGGAGCCCTACCACTACCTGCAGTTCGCTACTTCAAGATCAACAAGTTGGAG	480
Qy	161	LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln	180
Db	481	AAAGCTGTTCCTCGAGCACACACTTCTTCGTGGGCAATCCTGAGCACATGGAAATGCGAG	540
Qy	181	GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu	200
Db	541	CAGAACCTAGACTATTACCAACCATCTCTGGAGTGAAGGAGGCCGACTTCAAGGATCTT	600
Qy	201	GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln	220
Db	601	GAGACTCAACCCCATATGCAAGAAATTTGAGCTGGAGTGCAGTCTACTCAGAGGAACAG	660
Qy	221	ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu	240
Db	661	CCACAGGAAGCTGTGCCCCACCTAGAGCGCGCGTCCAGAAATACTTTGTGGCTATGAG	720
Qy	241	GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr	260
Db	721	GAGTGCCTGCGCTCTCGAAGGCGCCATGACTACGATCGCTACAACTACCTTGAGTAC	780
Qy	261	AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln	280
Db	781	AACGCTGACCTTCTCCAGGCCATCACAGATCATTCATCCAGGTCTCACTGTAGCAG	840
Qy	281	AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu	300
Db	841	AACGTGTCCGAGCTGTGCTTCCACCCCAAGTCGAGAGAAGCCCTTTGAAGACTTCCCTC	900
Qy	301	ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla	320
Db	901	CCATCGCATATATATATCTGCAGTTTGCTTACTATTAAC-----	939
Qy	321	GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn	340
Db	940	-----AAGACA-----AtCTGCTATTGTAAT960	960
Qy	341	LeuAlaTyrTyrAlaAlaMetLeuGlyGluHisThrArgSerIleGlyProArgGlu	360
Db	961	CTTCCTTGTCTCTCAAAATCTATAGAAA-----AAGAAG996	996
Qy	361	SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr	380
Db	997	AGTCCCAAGAGATACCGACACGAGCCTACTGGAAAAAGAACTGCTTTTCTTCGCTTAT	1056
Qy	381	AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro	400
Db	1057	GATGTTTTTGGAAATCCCTTTGTGGATCCGGATTCATGGACTCCAGAAAGATGATTCCTC	1116
Qy	401	LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu	420
Db	1117	AAGAGATTGCAAGAAACAGAGTCAAGACGGGAAACAGCCGTACGCACTCTCCAGGAG	1176
Qy	421	IleGlyAsnLeuMetLysGluIleGluThrIleuValGluGluLysThrLysGluSerIleu	440
Db	1177	ATTGGGAACCTTATGAAGAAATCGAGACCTTGTGGAAAGAAAGAACAGGAGTCACTG	1236
Qy	441	AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr	460
Db	1237	GATGTGACGACACTGACCCGGGAGGTGGCCCCCTGCTGTATGAAGGCATCAGTCTCAC	1296
Qy	461	MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp	480

Db	1297	ATGAATCTCCAAATCTCTGAATGGTTCCAGCGGGTGGTGAATGGACGGCGTAATCTCTGAC	1356
Qy	481	HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr	500
Db	1357	CACGAGTGTCCAGAGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGAGATGGCTAC	1416
Qy	501	ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys	520
Db	1417	CGGGGTGAGACCTCCCCACATATCCCAATGAAGAAGTTCTATGGTGCTACTGCTCTCAA	1476
Qy	521	AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn	540
Db	1477	GCCTCAAGCTGGGCAAGAAGCAAGTTCCTCTCAGAGTCCCACTGTACTACTACAAC	1536
Qy	541	ValThrGluLysValArgArgGlieMetGluSerTyrPheArgLeuAspThrProLeuTyr	560
Db	1537	GTACGGAGAAGTGGCGGCATCATGGAGTCTCTACTTCGCCCTGGATACGCCCTCTAC	1596
Qy	561	PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys	580
Db	1597	TTTTCTCTACTCTCATCTGGTGTGGCCTACCTCGACAGAGTCCAGGCAGAGAGGAAG	1656
Qy	581	AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrIleuValCys	600
Db	1657	GATGATAGTATCATCAGTCCACGTGGCAACTGATCTCTGAATCCGAGACCCCTCGTGTGT	1716
Qy	601	ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly	620
Db	1717	GTCAAGAGAGCCCCAGCGCTACACCTTCGCGACTACAGCGCCATCTCTTAAATGGG	1776
Qy	621	AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu	640
Db	1777	GACTTTCGATGCGGAATCTTTTATTTCACTGAATCGGATCCCAAGACCGTGACGGCAGAG	1836
Qy	641	ValGlnProGlnCysGlyArgAlaValAlaGlyPheSerSerGlyThrGluAsnProHisGly	660
Db	1837	GTGCAGGCTCAGTGTGGAAAGAGCGTGGGATCTCTTCAGGCATGAAACCCACATGGA	1896
Qy	661	ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro	680
Db	1897	GTGAAGGCTGTCAACAGGGGCGACGGCTGTGCCATCGCCCTGTGGTTTCACCCCTGGACCC	1956
Qy	681	ArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMetLeuPheSerPro	700
Db	1957	CGACACAGCGAGCGGGACAGGTCGACGCAGATGACCTGTGTAAGATGCTCTTCAGGCCA	2016
Qy	701	GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro	720
Db	2017	GAAGAGATGACCTCTCCAGGACGCCCTCGATCGCCAGAGGGTCCCCCGGAACT	2076
Qy	721	AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu	736
Db	2077	GCACAAGAGTCTCTCAGGCAGTGAATCGAAGGCCCAAGGATGAGCTA	2124
RESULT 10			
US-10-045-815-5			
; Sequence 5, Application US/10045815			
; Publication No. US20020160498A1			
; GENERAL INFORMATION:			
; APPLICANT: Wadhwa, Renu			
; APPLICANT: Sugihara, Takashi			
; APPLICANT: Ohide, Akiko			
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE			
; FILE REFERENCE: 06501-091001			
; CURRENT APPLICATION NUMBER: US/10/045,815			
; CURRENT FILING DATE: 2001-10-26			
; PRIOR APPLICATION NUMBER: PCT/JF00/02731			
; PRIOR FILING DATE: 2000-04-26			
; PRIOR APPLICATION NUMBER: JP 11/118806			
; PRIOR FILING DATE: 1999-04-26			
; NUMBER OF SEQ ID NOS: 15			
; SOFTWARE: FastSeq for Windows Version 4.0			

SECRET 10

```

; Result 10
; US-10-045-815-5
; Sequence 5, Application US/10045815
; Publication No. US20020160498A1
; GENERAL INFORMATION:
; APPLICANT: Wadhwa, Renu
; APPLICANT: Sugihara, Takashi
; APPLICANT: Oriide, Akiko
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; FILE REFERENCE: 06501-091001
; CURRENT APPLICATION NUMBER: US/10/045,815
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: PCT/JP00/02731
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: JP 11/118806
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0

```

; SEQ ID NO 5  
 ; LENGTH: 2416  
 ; TYPE: DNA  
 ; ORGANISM: Mus musculus  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (12)...(2252)  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: 2376  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-10-045-815-5

## Alignment Scores:

Pred. No.: 0 Length: 2416  
 Score: 3323.00 Matches: 630  
 Percent Similarity: 92.37% Conservative: 36  
 Best Local Similarity: 87.38% Mismatches: 52  
 Query Match: 85.87% Indels: 4  
 DB: 14 Gaps: 2

US-10-045-815-4 (1-736) x US-10-045-815-5 (1-2416)

QY 11 ThrLeuLeuAlaValAlaAlaAlaSer---GlnAlaGluValGluSerGluAlaGly 29  
 DB 40 ACGATGCTAGCGGTCGCGCGCGCGCGCGCTTACGGGTTCGGC-GAGTCTGAGCGCGGA 98  
 QY 30 TrpGlyMetValThrProAspLeuLeuPheAlaGluGlyThrAlaAlaValAlaArgGly 49  
 DB 99 TGGAGCTGGGAGCCCTTACCTCTTACGACAGGGGACCGGGGCTTACTCGCGCAGG 158  
 QY 50 AspTrpProGlyValValLeuSerMetGluArgAlaLeuArgSerArgAlaAlaLeuArg 69  
 DB 159 GACTGGCCCGGGTGGTCTTGAACATGAGCGGCTCTGGCTTCGCGGGCGGCGCTTGGCT 218  
 QY 70 AlaLeuArgLeuArgCysArgThrGlnCysAlaAlaAspPheProTrpGluLeuAspPro 89  
 DB 219 GCCCTCGCGCTGCGCTGCGCGCACACGCTGTGCCACCAACTGCGGTGGGACCGGACCTG 278  
 QY 90 AspTrpSerProSerPro-----AlaGlnAlaSerGlyAlaGlyAlaLeuArgAspLeu 107  
 DB 279 GATCTCGTCCGACCCCGCTGAGCGAGGACCGCGCGCGCGCGCTGCGACGACCTG 338  
 QY 108 SerPhePheGlyGlyLeuLeuArgAlaAlaCysLeuArgArgCysLeuGlyProPro 127  
 DB 339 CGCTTCTTCGAGCGCTGCTGCGCGCTGCGCGCTGCGCTGCGCTGCGCGCGCGCG 398  
 QY 128 AlaAlaHisSerLeuSerGluGluMetGluLeuGluPheArgLysArgSerProTyrAsn 147  
 DB 399 TCCTCCCACTTCTGAGTGAAGAACTCGACCTGAGGTTCAACAGCGGAGCGCGCTACAAC 458  
 QY 148 TyrLeuGlnValAlaLafyrPheLysIleAsnLysLeuGluLysAlaValAlaAlaHis 167  
 DB 459 TACCTGAGGTGCGCTATTCAGATATAACAAGCTGGAGAAAGCTGTGGCTGGCGCACAC 518  
 QY 168 ThrPhePheValGluAsnProGluHisMetGluMetGlnGlnAsnLeuAspTyrTyrGln 187  
 DB 519 ACCTTCTTTGGGCAATCTCTGAGCATATGAGATGCGGAGAACTTCGACTATTACCAA 578  
 QY 188 ThrMetSerGlyValLysGluAlaAspPheLysAspLeuGluThrGlnProHisMetGln 207  
 DB 579 ACCATGCTCGGGTGAAGGAGGAGAGCTTCAGGGATCTCGAGGCCAAGCCCATATGCA 638  
 QY 208 GluPheArgLeuGlyValArgLeuTyrSerGluGluGlnProGlnGluAlaValProHis 227  
 DB 639 GAGTTTCGGCTGGGGTACGACTTACTCAGAGGAGAACCCAGAGAGCTGTGGCCCCAC 698  
 QY 228 LeuGluAlaAlaLeuGlnGluTyrPheValAlaLafyrGluGluCysArgAlaLeuCysGlu 247  
 DB 699 CTGGAGCGGCACCTGCACAGAGTACTTGTGGCGGATGAGGAGTGGCTGCGCTCTGCGGA 758  
 QY 248 GlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAla 267

DB 759 GGCCCTATGACTAGCAGCGCTACAACTACTAGACTAGCGCTGACCTCTTCCAGGCC 818  
 QY 268 IleThrAspHisTyrIleGlnValLeuAsnCysLysGlnAsnCysValThrGluLeuAla 287  
 DB 819 ATCAGAGATCAATACGTCAGGTCTCACTCACTAGCAGAACTGTGTACGAGGTGGCT 878  
 QY 288 SerHisProSerArgGluLysProPheGluAspPheLeuProSerHisTyrAsnTyrLeu 307  
 DB 879 TCCACCCCAAGTAGGAAAGCCCTTGAAGACTTCTCCCTTCACACTATAAATACCTA 938  
 QY 308 GlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAlaGlyGluCysAlaLysThrTyr 327  
 DB 939 CAGTTTGCCTACTACAACTTGGAACTATACAAAGCTATTGAATGTGCCAAGACCTAC 998  
 QY 328 LeuLeuPhePheProAsnAspGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMet 347  
 DB 999 CTCCTCTTCTTCCCAATGATGAGGTGATGCACAGAACTCTGGCTATTACAGCCCATG 1058  
 QY 348 LeuGlyGluGluHisThrArgSerIleGlyProArgGluSerAlaLysGluTyrArgGln 367  
 DB 1059 CTTGGAGAAAGACAGCCAGCTCCATCAGCCCGGAGAAATGCCGAGAAATACCGAGCT 1118  
 QY 368 ArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyrAspValPheGlyIleProPhe 387  
 DB 1119 CCAAACTCTTTCGAGAAAGAACTGCTTTTCTTCGCTTATGACATTTTGAATTCCTTT 1178  
 QY 388 ValAspProAspSerTrpThrProGluGluValIleProLysArgLeuGlnGluLysGln 407  
 DB 1179 GTGGATCCGATTCATGAGCTCCAGAGAGATGATCCCAAGAGATTCCAGAGAAAGACAG 1238  
 QY 408 LysSerGluArgGluThrAlaValArgIleSerGlnGluIleGlyAsnLeuMetLysGlu 427  
 DB 1239 AAGTCTGAACGGAAACAGCCGCTGCTCCAGAGAGATTGGGAACCTTATGAGGAA 1298  
 QY 428 IleGluThrLeuValGluGluLysThrLysGluSerLeuAspValSerArgLeuThrArg 447  
 DB 1299 ATCGAGACCTTGTGGAAAGAGAACCAAGAGAGTCTCTGGATGTGAGCAGACTGACCGG 1358  
 QY 448 GluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThrMetAsnSerLysLeuLeuAsn 467  
 DB 1359 GAAGTGTGCTCTGCTGTATGAGGCACTAGTCTCCATGAACTCCAAAGTCTTGAAT 1418  
 QY 468 GlyTyrGlnArgValValMetAspGlyValIleSerAspHisGluCysGlnGluGln 487  
 DB 1419 GGCTCCAGCGGTGCTGATGGTGTGATCTCTGATGATGATGATGATGATGATGATGATG 1478  
 QY 488 ArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHis 507  
 DB 1479 AGACTGACCAATGCGGCGAGCAACTTCGGGAGATGGCTACCGAGGTGAGCTCCCCAC 1538  
 QY 508 ThrProAsnGluLysPheTyrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGlu 527  
 DB 1539 ACCCCAAATGAAAGTTCTATGGTGTACTGTCTCTCAAGCTCTCAAGCTCGGCGAGAA 1598  
 QY 528 GlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsnValThrGluLysValArgArg 547  
 DB 1599 GGAAAGTCTCTGCGAGAGTCCCGGATGACTACCAACGTCGACAGAGAGAGGTGCGCGC 1658  
 QY 548 IleMetGluSerTyrPheArgLeuAspThrProLeuTyrPheSerTyrSerHisLeuVal 567  
 DB 1659 GTCTAGGAGTCTTCTCCGCTCGAGACCGCTCTATTCTCTTCTTCTTCTTCTTCTTCT 1718  
 QY 568 CysArgThrAlaIleGluGluValGlnAlaGluArgLysAspSerHisProValHis 587  
 DB 1719 TGCCGCACTGCATAGAGAGTTCACGGCTGAGAGAGAGAGAGTAGTAGCCCGCTCCAC 1778  
 QY 588 ValAspAsnCysIleLeuAsnAlaGluThrLeuValCysValLysGluProProAlaTyr 607  
 DB 1779 GTGGATACTGCATCTCTGATGCGGAGGCTTCTGATGTGATGATGATGATGATGATGATG 1838  
 QY 608 ThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPhe 627  
 DB 1839 ACCTTCCGGGAATACAGCGCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1898

```
QY 628 TyrPheThrGluLeuAspAlaValThrAlaGluValGlnProGlnCysGlyArg 647
Db 1899 TACTTCCACAGAACTAGATCCCAAGACTGTGACGGCAGAGGTGCACCCCGAGTGTGGAAGG 1958
QY 648 AlaValGlyPheSerSerGlyThrGluAsnProHisGlyValIysAlaValThrArgGly 667
Db 1959 GCTGTGGGAATTCCTTCTGGCACTGAGAACCCACATGAGAGTGTACACAGGGGG 2018
QY 668 GlnArgCysAlaIleAlaLeuTrpPheThrLeuAspProArgHisSerGluArgAspArg 687
Db 2019 CAGCGCTGGCCATCGCCCTGTGTTTACGCTGATCTCGGCACACAGTGCAGAGACAGG 2078
QY 688 ValGlnAlaAspAspLeuValIysMetLeuPheSerProGluMetAspLeuSerGln 707
Db 2079 GTGCAGGCAGATGACCTGTGTGAAGATCTGTTCAGCCCAAGAGGTGACCTCCCCAG 2138
QY 708 GluGlnProLeuAspAlaGlnGlnGlyProProGluProAlaGlnSerLeuSerGly 727
Db 2139 GAACGCCCTTGCCTGACCCAGCAGGCTTCGCCAGAGCCTGGAGAGATTTCTGCATGCT 2198
QY 728 Ser 728
Db 2199 GCT 2201

RESULT 11
US-10-045-815-7
; Sequence 7, Application US/10045815
; Publication No. US20020160498A1
; GENERAL INFORMATION:
; APPLICANT: Wadhwa, Renu
; APPLICANT: Sugihara, Takashi
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; FILE REFERENCE: 06501-091001
; CURRENT APPLICATION NUMBER: US/10/045, 815
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: PCT/JP00/02731
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: JP 11/118806
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 2322
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12)...(1637)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 2282_
; OTHER INFORMATION: n = A,T,C or G
US-10-045-815-7

Alignment Scores:
Pred. No.: 0 Length: 2322
Score: 3120.00 Matches: 602
Percent Similarity: 87.97% Conservative: 34
Best Local Similarity: 83.26% Mismatches: 50
Query Match: 80.62% Indels: 38
DB: 14 Gaps: 3

US-10-045-815-4 (1-736) x US-10-045-815-7 (1-2322)
QY 11 ThrLeuLeuAlaValAlaAlaAspSer--GlnAlaGluValGluSerGluAlaGly 29
Db 40 ACATGTACGGTTCGGCGCGCGCGCGCTTACGGGTTCCGGC-GAGTCTGAGCGGGGA 98
QY 30 TrpGlyMetValThrProAspLeuPheAlaGluGlyThrAlaAlaValThrAlaArgGly 49
Db 99 TGGACGTGGCGAGCCCTGACCTGCTTACGACAGGGGACCGCGGCTACTCGCGCAGG 158
```

```
QY 50 AspTrpProGlyValValLeuSerMetGluArgAlaLeuArgSerArgAlaAlaLeuArg 69
Db 159 GACTGGCCCGGGGTGGTCTGAACATGAGCGGGCTCTGGCTCGCGGGGGCCCTCGCT 218
QY 70 AlaLeuArgLeuArgCysArgThrGlnCysAlaAlaAspPheProTrpGluLeuAspPro 89
Db 219 GCCCTCGCGCTCGCTCGGCACACAGCTGTGCCACCGAAGTGCCTGGGACCGGACCTG 278
QY 90 AspTrpSerProSerPro-----AlaGlnAlaSerGlyAlaGlyAlaLeuArgAspLeu 107
Db 279 GATCTCGTCCGGACCCCGAGCTGAGCAGGAGACCCGGCGCGCGCCCTGCGACGAGCTG 338
QY 108 SerPhePheGlyGlyLeuLeuArgAlaAlaCysLeuArgArgCysLeuGlyProPro 127
Db 339 CGCTTCTCGAGCGCGTCTCGCGCTGCGCTGCGCTGCGCTGCGTGGCGCGCGCC 398
QY 128 AlaAlaHisSerLeuSerGluGluMetGluLeuGluPheArgIysArgSerProTyrAsn 147
Db 399 TCTGCCCACTTGTCTGAGTGAAGAACTGGACCTGTGAGTTCAACAAGCGAGGAGCCGTACAAC 458
QY 148 TyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGluLysAlaValAlaAlaHis 167
Db 459 TACCTGACAGTGGCTTATTTCAAGATAAACAAGCTGGAGAAAGCTGTGGCTGGCGCAC 518
QY 168 ThrPhePheValGlyAsnProGluHisMetGluMetGlnGlnAsnLeuAspTyrTyrGln 187
Db 519 ACCTTCTTTGTGGCAATCTCTGAGCACATGGAGATGGCGAGAACTCGACTATTACCAA 578
QY 188 ThrMetSerGlyValLysGluAlaAspPheLysAspLeuGluThrGlnProHisMetGln 207
Db 579 ACCATGCTGGGGTGAAGGAGGAGACTTCAGGAGATCTCGAGGCAAGCCCATATGAT 638
QY 208 GluPheArgLeuGlyValArgLeuTyrSerGluGlnGlnProGlnGluAlaValProHis 227
Db 639 GAGTTTCTGGCTGGGGTACGACTCTACTCAGAGGAGAGCCACAGGAAGCTGTGCCCCAC 698
QY 228 LeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGluGluCysArgAlaLeuCysGlu 247
Db 699 CTGGAGCGCGCACTGCAAGAGTACTTTGTGGCCGATGAGAGTGGCGGCTCTGCCAA 758
QY 248 GlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAla 267
Db 759 GGGCCCTATGACTACGACGGCTCAACTCTAGACTACAGCGGTGACCTCTTCCAGGCC 818
QY 268 IleThrAspHisTyrIleGlnValLeuAsnCysLysGlnAsnCysValThrGluLeuAla 287
Db 819 ATCACAGATCATTACGTCAGGTCCTCACTGTAAGCAGAACTGTGTCCGAGGTGGCT 878
QY 288 SerHisProSerArgGluLysProPheGluAspPheLeuProSerHisTyrAsnTyrLeu 307
Db 879 TCCCAACCAAGTAGGAAAGCCCTTTGAAGACTTCTCCCTTACACTATATATACCTA 938
QY 308 GlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAlaGlyGluCysAlaLysThrTyr 327
Db 939 CAGTTTGGCTACTACAACTTGGGAACCTATACACAAGCTATTGAATGTGCCAAGACCTAC 998
QY 328 LeuLeuPhePheProAsnAspGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMet 347
Db 999 CTCTCTTCTTCTTCCCAATGATGAGGTGATGCACCAAGATCTGGCTTATTACACAGCCATG 1058
QY 348 LeuGlyGluGluHisThrArgSerIleGlyProArgGluSerAlaLysGluTyrArgGln 367
Db 1059 CTTGGAGAGAGAGAGCGCCAGCTCCATCAGCCCGGAGAGATGCCGAGGAATACCGAGCT 1118
QY 368 ArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyrAspValPheGlyIleProPhe 387
Db 1119 CCAAAACCTGTTGGAGAAAGAACTGCTTTTCTTCGCTTATGACATTTTGGAAATTCCTTT 1178
QY 388 ValAspProAspSerTrpThrProGluGluValIleProLysArgLeuGlnGluLysGln 407
Db 1179 GTGGATCCCGATTATGACTCCAGAGAGATGATTTCCAGAGAGATTCCAGAGAGAAACAG 1238
```

```
QY 408 LysSerGluArgGluThrAlaValArgIleSerGlnGluIleGlyAsnLeuMetLysGlu 427
DB 1239 AAGTCTGAACGGGAAACAGCGGTACGATCTCCAGAGATGGGAACCTTATGAGGAA 1298
QY 428 IleGluThrLeuValGluGluLysThrLysGluSerLeuAspValSerArgLeuThrArg 447
DB 1299 ATCCAGACCCCTGTGGAAAGACAGAACCAAGAGTCTCTGGATGTGAGCAGACTGACCCGG 1358
QY 448 GluGlyGlyProLeuLeuLeuThrGlyIleSerLeuThrMetAsnSerLysLeuLeuAsn 467
DB 1359 GAAGTGTGTCCTCTGTATGAGGACATAGTCTCACCATGAACCTCAAGTCTTGAAT 1418
QY 468 GlyTyrGlnArgValMetAspGlyValIleSerAspHisGluCysGlnGluLeuGln 487
DB 1419 GGCTCCACGGGTGTGATGTGATGTTGTATCTCTGATGATGATGCCAGGAGTGCAG 1478
QY 488 ArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyrArgGlyGlnThrSerProHis 507
DB 1479 AGACTGACCAATGCGGCAGCAACTTCGGAGATGCTACCGAGGTGAGCTCCCCAC 1538
QY 508 ThrProAsnGluLysPheTyrGlyValThrValPheLysAlaLeuLysLeuGlyGlnGlu 527
DB 1539 ACCCCAAATGAAAGTCTATGTTGTTACTGTCTCAAGCTCTCAAGCTCGGGCAGGAA 1598
QY 528 GlyLysValProLeuGlnSerAla-HisLeuTyrTyrAsnValThrGluLysValArg 547
DB 1599 GGAAGAGTTCCTGCGCAGAGTGCCTGCAC----- 1627
QY 547 rglIleMetGluSerTyrPheArgLeuAspThrProLeuTyrPheSerTyrSerHisLeuV 567
DB 1627 ----- 1627
QY 567 alCysArgThrAlaIleGluValGlnAlaGluArgLysAspAspSerHisProValH 587
DB 1628 -----CGCACTGCAATAGAGAGTCACAGGCTGAGAGGAGGACAGTAGCCACCCCGTCC 1682
QY 587 isValAspAsnCysIleLeuAsnAlaGluThrLeuValCysValLysGluProProAlaT 607
DB 1683 ACGTGGATAACTGTCATCTGATGCGGAGCTTGTATGATCAAGAGGACCCCGACAT 1742
QY 607 yrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsn 627
DB 1743 ACAGTTCGCGGAATACAGCGCCATCTTTACCTCAATGCGGACTTCGATGGAGGAACT 1802
QY 527 heTyrPheThrGluLeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyA 647
DB 1803 TTTACTTTCACAGAACTAGATGCCAAGACTGTGACGCGCAGAGGTGCAGCCCGTGTGGAA 1862
QY 647 rGAlaValGlyPheSerSerGlyThrGluAsnProHisGlyValLysAlaValThrArg 667
DB 1863 GGGCTGTGGGATCTCTTCTTGGCACTGAGAACCCACATGAGAGTGAAGGCTGTCCACAGG 1922
QY 667 lyGlnArgCysAlaIleAlaLeuThrPheThrLeuAspProArgHisSerGluArgAsp 687
DB 1923 GCGAGCGCTCGCCATCGCCCTGTGGTTACGCTGGATCTCGGCACAGTGCAGAGAGACA 1982
QY 687 rGValGlnAlaAspAspLeuValIysMetLeuPheSerProGluGluMetAspLeuSerG 707
DB 1983 GGGTGCAGGAGATGACCTGTGAGATGCTGTTCAGCCAGAGAGGTGGACCTCCCCC 2042
QY 707 lnGluGlnProLeuAspAlaGlnGlnGlyProGluProAlaGlnGluSerLeuSerG 727
DB 2043 AGCAACAGCCCTCGCTGACACAGCGGTTCGCCAGAGCTGCGAAGAGATTTCGATG 2102
QY 727 lySer 728
DB 2103 GTGCT 2107
```

RESULT 12

US-10-302-172-226

; Sequence 226, Application US/10302172

; Publication No. US20040053250A1

; GENERAL INFORMATION:

```
; APPLICANT: Tang, Y. Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: NO. US20040053250A1el Arginine-rich Protein-like Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 803.1CNC
; CURRENT APPLICATION NUMBER: US/10/302,172
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/225,251
; PRIOR FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: PCT US02/05095
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 09/799,451
; PRIOR FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 950
; SOFTWARE: PC_FL_Genes Version 2.0
; SEQ ID NO 226
; LENGTH: 1437
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (32)..(1255)
; US-10-302-172-226
Alignment Scores:
Pred. No.: 4,526-229 Length: 1437
Score: 2147.00 Matches: 406
Percent Similarity: 99.51% Conservations: 0
Best Local Similarity: 99.51% Mismatches: 2
Query Match: 55.48% Indels: 0
DB: 13 Gaps: 0
US-10-045-815-4 (1-736) x US-10-302-172-226 (1-1437)
```

```
QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValValAlaAlaSer 20
DB 32 ATGGCGGTACCGCGCTTGAAGCTGTACCACACTGTGGCTGTGGCGCGCTGCCTCC 91
QY 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla 40
DB 92 CAAGCCGAGTCCGAGTCCGAGCGAGGATGGGCACTGGTACGCTGTCTCTCTGCC 151
QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60
DB 152 GAGGGGACCGCAGCCTACGCGCGGGGACTGGCCCGGGTGGTCTGAGCATGGAACGG 211
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgCysArgThrGlnCysAla 80
DB 212 GCGCTGCGCTCCCGGCGAGCCCTCGGCGCCCTTCGCTGCGCTGCGCGACCCAGTGTGCC 271
QY 81 AlaAspPheProTyrGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100
DB 272 GCGCACTTCCCGTGGAGCTGGACCCGCACTGTGTCCCGCGCGCGCGCGCGCGCGCC 331
QY 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyLeuLeuArgArgAlaAlaCysLeu 120
DB 332 GCGCGCGCGCTGCGCGAGCTTCTTCTGGGGGCGCTTCTGCGCTGCGCTGCGCTGCGCTG 391
QY 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe 140
DB 392 GCGCGCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 451
QY 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160
DB 452 CGCAAGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 511
QY 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180
DB 512 AAAGCTGTGCTGCGAGCACACACCTTCTTCTGTGGGCACTCTGAGCACATGGAATGAG 571
QY 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200
```

572 CAGAACTAGACTATTACCAACCATGCTGGAGTGAAGAGCCGACTTCAAGGATCTT 631  
QY 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220  
Db 632 GAGACTCAACCCCATATGCAAGAAATTCGACTGGAGTGCAGCTCTACTACAGAGAACAG 691  
QY 221 ProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
Db 692 CCACAGGAAGCTGCCCCACCTAGAGCGCGCTCAAGAAATCTTTGGCCCTATGAG 751  
QY 241 GluCysArgAlaLeuGluGluGluGluGluGluGluGluGluGluGluGluGlu 260  
Db 752 GAGTGGCTGCGCTCTGGAGGGCCCTATGACTACGATGGCTACAACTACCTTGAGTAC 811  
QY 261 AsnAlaAspLeuPheGlnAlaLeuThrAspHisTyrIleGlnValLeuAsnGlyGln 280  
Db 812 AAGCTGACCTCTTCCAGGCGCATCAGAGTCAATATACATCCAGGTCTCTCACTGTAAGCAG 871  
QY 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluGluGluGluGluGluGlu 300  
Db 872 AACGTGTCTCAGGAGCTTGTCTTCCACCAGTCGAGAGAGCCCTTTGAGACTTCTTC 931  
QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
Db 932 CCATCGCATTAATATATCTGAGTTTGCCTACTATAAATTCGGAATATATACACAGCT 991  
QY 321 GlyGluCysAlaTyrThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGln 340  
Db 992 GTTGAATGTGCCAGACTATCTCTCTTCTCCCAATGACGAGGTGATGAACCAAAAT 1051  
QY 341 LeuAlaTyrTyrAlaAlaMetLeuGluGluGluHisThrArgSerIleGlyProArgGlu 360  
Db 1052 TTGGCTTATATGAGTATGCTTGGAGAGACACACACAGATCCATCGGCCCGCTGAG 1111  
QY 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluGluGluGluGluGluGluGlu 380  
Db 1112 AGTCCCAAGGAGTACCCAGCAGCAAGCTACTAGTGAAGAGACACACAGATCCATCGGCCCGCTGAG 1171  
QY 381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400  
Db 1172 GATGTTTTTGAATTCCTTTTGTGGATCCGGATTCATGGACTCCAGAGAGATGATTCCTCC 1231  
QY 401 LysArgLeuGlnGluLysGlnLys 408  
Db 1232 AAGAGATTGCAAGAGAAACAGAG 1255

## RESULT 13

US-10-094-749-921  
; Sequence 921, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEKI, NAOHICO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTOUKI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA  
; FILE REFERENCE: 084335/0160  
; CURRENT APPLICATION NUMBER: US/10/094,749  
; CURRENT FILING DATE: 2002-03-12

; PRIOR APPLICATION NUMBER: 60/350,435  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: JP 2001-328381  
; PRIOR FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 3381  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 921  
; LENGTH: 3396  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-10-094-749-921  
  
Alignment Scores:  
Pred. No.: 2,88e-184 Length: 3396  
Score: 1751.00 Matches: 356  
Percent Similarity: 63.52% Conservative: 102  
Best Local Similarity: 49.38% Mismatches: 211  
Query Match: 45.25% Indels: 52  
DB: 16 Gaps: 12  
  
US-10-045-815-4 (1-736) x US-10-094-749-921 (1-3396)  
QY 23 GluValGluSerGluAlaGlyTrpGlyMetValThrPro---AspLeuLeuPheAlaGlu 41  
Db 295 GAGCTGGAGCTGGAGCCC-----GGGCTCTGCGAGCCCTTCGACCTCTCTACCCGACG 348  
QY 42 GlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluAla 61  
Db 349 GCGCGGCGCGCTACTACAGCGGAGACTACGAGCGCGGTGCGGACTTGGAAAGCGCG 408  
QY 62 LeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAlaAla 81  
Db 409 CTGCGCAGCACCGCGCGCTGCGGAAATCCGACGCGCTGTGCCCGCCACTGCGCGCGCG 468  
QY 82 AspPheProTrpGluLeuAspProAspTrpSerProAlaGlnAlaSerGlyAla 101  
Db 469 CGCCACCCG-----CTCCGCGCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCG 507  
QY 102 GlyAlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgArgAlaAlaCysLeuArg 121  
Db 508 GCGGCT-----GAGCTGCGCTCTTTCGCTCTTGTGGCGCGCGCGCTGTATCGC 561  
QY 122 ArgCys-----LeuGlyProProAlaAla---HisSerLeuSerGluGluMet 136  
Db 562 AGCTGTGAGACCCAGCGCTCGGGGCGCCCGCATCCGCCACCGGTACGCGAGGATGTG 621  
QY 137 GluLeuGluPheArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIle 156  
Db 622 CGCAGCGACTTCCAGCGCAGAGTGCCTACAACTACCTGCGCGCGCGCTTACATCAAGCTT 681  
QY 157 AsnLysLeuGluLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHis 176  
Db 682 AACACGCTCGAAAGACAGTGAAGCAGCTACACATTTTCGTGGCTAACCTGAGCAC 741  
QY 177 MetGluMetGlnGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAsp 196  
Db 742 ATGGAATGCGACGAGAACATTGAGATTACAGCGCGCAGCTGGTGTGAAGCATTCGAC 801  
QY 197 PheLysAspLeuGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyr 216  
Db 802 TTGCTAGACAGAGAGCCAGCCACATGAGGAGTTTACAATCGAGAGCTTAAACATTTAT 861  
QY 217 SerGluGluGlnProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPhe 236  
Db 862 GAGGCTGAGCTTTGAGATGGCTATCAGGCATTTGGAACAAGCTTTAGAGAAATATTTC 921  
QY 237 ValAlaTyrGluGluCysArgAlaLeuGluGluGluGluGluGluGluGluGluGlu 256  
Db 922 GTTGAAGATACAGAAATGCCGCGACCTTATGTGAGGGGCGCTCAGAGATTTGAAGAAATATGAG 981  
QY 257 TyrLeuGluTyrAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeu 276  
Db 982 TATTTAGGGTATAAGGCTGCTCTGTATGAAGCTATTTCGAGATCCTACATCAGCGGTCTT 1041



QY 277 AsnCysLysGlnAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPhe 296  
 Db 1042 GTTGTGACGATGATGTTGAGGAGACTTGCACCCCGCTTGGCGGCTCTCTCCCATC 1101  
 QY 297 GluAspPheLeuProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsn 316  
 Db 1102 GAGAAATTTCTCTCTCTGCACTATGATTACCTACAGTTTGCTACTATCGAGTTGGTGG 1161  
 QY 317 TyrThrGlnAlaGlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluVal 336  
 Db 1162 TATGTGAAGCCCTCGAGTGTGCCAAGCCCTATCTTATGCCATCCAGATGATGAGGAT 1221  
 QY 337 MetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGluGlu--HisThrArgSer 355  
 Db 1222 GTCTTAGACAATGTGATTACTATGAGAGTCTGTGGATGATAGCATTTGACCGGCATCC 1281  
 QY 356 IleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeu 375  
 Db 1282 ATTGAGGCCAGAGAGGATTTAACAAATGTTTGTGAACGTCATAAGCTGAGTCTGAGCTG 1341  
 QY 376 LeuPhePheAlaTyrAspValPheGlyIleProPheValAspProAspSerTyrThrPro 395  
 Db 1342 ATAAATCAGTGCAGAGGTCTGGGTTTTCATACACTGAACCAATTTATGATC--- 1398  
 QY 396 GluGluValIleProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaVal 415  
 Db 1399 -----AGATATGAGGACGACAGGATGAGAAT 1425  
 QY 416 ArgIleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLys 435  
 Db 1426 CGGGCCCTTCAGGATGAACGTAAGAGGACAGAGTTCATGATTCTCAATGGGAAAA 1485  
 QY 436 ThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGlu 455  
 Db 1486 ---AAGCTATCACCAAGATAGTCGACCTAAGAGAGGTGCTCTCTACTCTATGAG 1542  
 QY 456 GlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAsp 475  
 Db 1543 AACATCATCTGCTCACTCAACTCGAGCAGCTGAACGGGACTCAGCGGTTCTCTCGGAT 1602  
 QY 476 GlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThr 495  
 Db 1603 AACCTCTGTGGAAGAACAGTGCAGAGACTCCACAGCTGGCCAGTGAATCATGCTT 1662  
 QY 496 SerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGly 515  
 Db 1663 GTTGTGATGGATACAGAGGAAAACTTCAACCCATACACCAATGAAAGTTTGAAGGT 1722  
 QY 516 ValThrValPheLysAlaLeuLysLeuGluGlnGlyLysValProLeuGlnSerAla 535  
 Db 1723 GCACTGTCTGAAAGCACTCAATCTGTTATGAAGTCTGATCCACATGAAGAGCGCT 1782  
 QY 536 HisLeuTyrTyrAsnValThrGluLysValArgIleMetGluSerTyrPheArgLeu 555  
 Db 1783 CGTCTGTTTATGATACAGCGAAAGCTCGAAGGATTGTAGAACTTATTTATGCTG 1842  
 QY 556 AspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluVal 575  
 Db 1843 AACTCAACTCTGTTATTTTCTATACACATGCTGTGCGCAACAGCCCTGTCTGTCTAG 1902  
 QY 576 GlnAlaGluArgLysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAla 595  
 Db 1903 CAGGATAGAAGAAATGACCTCAGTCATCCATCCATGCTGACAACTGTTTGTGGATCCA 1962  
 QY 596 GluThrLeuValCysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIle 615  
 Db 1963 GAGGCCAACGAATGCTGGAGAGGAGCCCTCTGCTTACACATTTTCAGACTATAGTCTCTC 2022  
 QY 616 LeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLys 635  
 Db 2023 CTATATATGATGATGATCTTTGAAGGAGGAGAAATTCATATTTCAGAGATGATGCTAG 2082

QY 636 ThrValThrAlaGluValGlnProGlnCysGlyArgAlaValAlaGlyPheSerSerGlyThr 655  
 Db 2083 ACTGTGACTGCTCTTATAAAACCAAAATGTGGCGCATGATCAGCTTCTCATCTGGAGA 2142  
 QY 656 GluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyr 675  
 Db 2143 GAGAACCCTCATGGGTGAAGCAGTCCACCAAGGGAAGAGGTGTGCTGTGCTCTGTGG 2202  
 QY 676 PheThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspLeuValLys 695  
 Db 2203 TTCACCTTGGACCCACTTTATAGAGAAATTTGGAGCAATACAGGCTGATGAAGTATGCA 2262  
 QY 696 MetLeuPheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGln 715  
 Db 2263 ATT-----CTGGATCAAGAA-----CAGCAA 2283  
 QY 716 GlyProProGluProAlaGlnGluSerLeuSerGlySerGluSerLysLysAspGlu 735  
 Db 2284 GCGAAGCATGAA-----CTGAATATCAACCCCTAAAGATGAG 2319  
 QY 736 Leu 736  
 Db 2320 CTA 2322

RESULT 14  
 US-10-071-766-86  
 ; Sequence 86, Application US/10071766  
 ; Publication No. US20020192678A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Huel-Wei Chen  
 ; TITLE OF INVENTION: GENES EXPRESSED IN SENESENCE  
 ; FILE REFERENCE: PA-0043 US  
 ; CURRENT APPLICATION NUMBER: US/10/071,766  
 ; CURRENT FILING DATE: 2002-02-07  
 ; NUMBER OF SEQ ID NOS: 144  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 86  
 ; LENGTH: 2665  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; OTHER INFORMATION: Inocyte ID No. US20020192678A1 978730.4  
 ; NAME/KEY: unsure  
 ; LOCATION: 353-387, 694, 727  
 ; OTHER INFORMATION: a, t, c, g, or other  
 US-10-071-766-86

Alignment Scores:  
 Pred. No.: 1,8e-180 Length: 2665  
 Score: 1715.50 Matches: 356  
 Percent Similarity: 63.49% Conservative: 103  
 Best Local Similarity: 49.24% Mismatches: 211  
 Query Match: 44.33% Indels: 53  
 DB: 14 Gaps: 12

US-10-045-815-4 (1-736) x US-10-071-766-86 (1-2665)

QY 23 GluValGluSerGluAlaGlyTyrGlyMetValThrPro---AspLeuLeuPheAlaGlu 41  
 Db 430 GAGCTGGAGCTGGAGCCC-----GGGCCTCTGCGACCTTCGACCTCTCTACGCCAGC 483  
 QY 42 -GlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArgAl 61  
 Db 484 GGGCGCGGCCCTCTACTACCGGAGACTACGACGAGCGGTGCGCGACTTGGAGCGGC 543  
 QY 61 aLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArg-ThrGlnCysAlaA 81  
 Db 544 GCTGCGAGCCACCGCGGCTTGGGAAATCCGACGCGCTGTGCCCTGCCACTGCCGCG 603  
 QY 81 laAspPheProTyrGluLeuAspProAspTyrPheProSerProAlaGlnAlaSerGlyA 101  
 Db 604 CGCGCACCCG-----TCTCCGCGCCCGCCCGCGCGGAGGGG 642







Db	1161	ATCCAGCGCTTCATCTCCGATCCCTGGGGGAGAGAGCAGCTCTACTATGCCATGGAG	1220
Qy	382	ValPheGlyIleProPheValAspProAspSerThrProGluGluValIleProLys	401
Db	1221	CACCTGGGACCAAGCTTCAAGATCCTGACCCCTGACCCCTGCAGCTCTCATCCTCGAG	1280
Qy	402	ArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGluIle	421
Db	1281	GCACCTTAGAGAAAGCTCAGAGAGGATCAGAGAAG-----	1316
Qy	422	GlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeuAsp	441
Db	1317	-----AGGCCTTGGACCATGAGCCCGTGAAG	1343
Qy	442	ValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrgluGlyIleSerLeuThrMet	461
Db	1344	CCAAAGCCCTTGACCTACTGGAAGGATGCTCTCTCTGAGGGTGTGACCTTGACCCAG	1403
Qy	462	AsnSerLysLeuLeuAsnGlyTyrginArgValValMetAspGlyValIleSerAspHis	481
Db	1404	GATTCAGGAGCTGATGGTGGAGCGGCGGTGTGGATGGGTGCTCACCCCCAGCC	1463
Qy	482	GluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGly-----AspGly	499
Db	1464	GAGTGTGGGTGTGCTGTCAGCTGGCTAAGGATGCAGCTGGGGCTGGAGCCAGGTCTGGC	1523
Qy	500	TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe	519
Db	1524	TATCGTGTGCGGCTCCCTCACACCCCATGAGCGCTTCAGGGGCTCACGGTGTCTT	1583
Qy	520	LysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyr	539
Db	1584	AAGCTCGCAGCTGGCGCGGTGGGACAGTGGGCAGTCAGGGTGTGAAGCTGCTTCTG	1643
Qy	540	AsnValThrGluLysValArgAlaGlyIleMetGluSerTyrPheArgLeuAspThrProLeu	559
Db	1644	GAGTGTAGCGGGGTGCGGACCTTACCCAGCCCTACTTCTCCCGGAGCGGCCCTG	1703
Qy	560	TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArg	579
Db	1704	CATCTGTCCTTACCCACCTGGTGTGCGCAGCGCCATAGAGGAGAGCAAGAGCAGCGC	1763
Qy	580	LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal	599
Db	1764	ATGACCTGTGATCACCAGTGCACGCAGACAACTGCGTCTGACCCCTGACACCGGGAGAG	1823
Qy	600	CysValLysGluProProAlaTyThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn	619
Db	1824	TGCTGGGGGAGCCCCAGCCTACACCTATCGGACTACAGCGGACTCTCTACCTCAAC	1883
Qy	620	GlyAspPheAspGlyGlyAsnPheTyThrGluLeuAspAlaLysThrValThrAla	639
Db	1884	GATGACTTCCAGGTGGGGACCTGTCTTACGGAGGCCAACGCCCTCACTGTCACGGCT	1943
Qy	640	GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis	659
Db	1944	CGGGTGGCTCTCGCTGTGGGGCGCTTGTGGCCCTTACGCTCCGCTGTCGAGATCCCAT	2003
Qy	660	GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuIrrpPheThrLeuAsp	679
Db	2004	GGGGTGTGGCGGTGACTCGGGGACGGCGCTGTGCCCTGACCTGTGGCACACGTGGGCA	2063
Qy	680	ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer	699
Db	2064	CCTGAGCACAGGAGCAGAGGTGATAGAGCAAGAAAGAACTGCTGCAG-----GAG	2114
Qy	700	ProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGlu	719
Db	2115	TCACAGGAGGAGGAGAGAGAGAGAAATGCCACAAAGACCTTCCCCAGAG	2174
Qy	720	ProAlaGln-----GluSerLeuSerGlySerGluSerLysProLys	733
Db	2175	CCCCCTAGCCGACGACAGAGGGTCCAGACAGACTGGAGAGGCGACCTCGGGTTCGG	2234

Qy 734 AspGluLeu 736  
Db 2235 GAGGAGCTG 2243

Search completed: July 18, 2004, 18:22:27  
Job time : 1163 secs

**This Page Blank (uspto)**

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 09:59:50 ; Search time 20888 Seconds

(without alignments)  
1527.215 Million cell updates/sec

Title: US-10-045-815-4

Perfect score: 3870

Sequence: 1 MAVRALKLTLLAVVAAS.....PPEPAQESLSGSESKPKDEL 736

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 segs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/US10045815/runat\_14072004\_123008\_16818/app\_query.fasta\_1.903  
-DB=GenEmbl -QPMF=fastap -SUFFIX=std.rge -MINMATCH=0.1 -LCORCL=0 -LCORCL=0  
-UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdd -LIST=45  
-DOCFMT=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10045815 @CGN 1 1 7406 @runat\_14072004\_123008\_16818 -NCFU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*

- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.nam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Query Length	DB ID	Description
1	3870	100.0	2600	9	AF097432 Homo sapi
2	3846	99.4	2563	6	AX136191 Sequence
3	3846	99.4	2563	6	BD123548 Secretary
4	3846	99.4	2563	9	AK075418 Homo sapi
5	3846	99.4	2583	6	AX704765 Sequence
6	3846	99.4	2585	6	AX877288 Sequence
7	3846	99.4	2585	6	BD156581 Primer fo
8	3846	99.4	2585	9	AK027697 Homo sapi
9	3844	99.3	2524	6	AX877139 Sequence
10	3844	99.3	2524	6	BD156505 Primer fo
11	3844	99.3	2524	9	AK027680 Homo sapi
12	3780	97.7	2829	9	AF097431 Homo sapi
13	3726.5	96.3	2393	6	AX880254 Sequence
14	3726.5	96.3	2393	6	BD158280 Primer fo
15	3726.5	96.3	2393	9	AK027648 Homo sapi
16	3403.5	87.9	2569	10	BC024047 Mus muscu
17	3383.5	87.4	2524	10	AF087433 Rattus no
18	3323	85.9	2391	10	AF165163 Mus muscu
19	3120	80.6	2297	10	AF165164 Mus muscu
20	2085	53.9	1562	9	BC015309 Homo sapi
21	2034	52.3	1173	9	BT007039 Homo sapi
22	2024	52.3	1173	12	BT007768 Synthetic
23	2022	52.2	1526	9	AK025841 Homo sapi
24	1751	45.2	3396	6	AX714237 Sequence
25	1751	45.2	3396	6	AK056447 Homo sapi
26	1747	45.1	2127	6	AX815645 Sequence
27	1747	45.1	2261	6	AX815648 Sequence
28	1747	45.1	3386	9	AJ430351 Homo sapi
29	1747	45.1	3496	9	AK125334 Homo sapi
30	1715.5	44.3	2249	10	MMU430350 Mus muscu
31	1575.5	40.7	2789	6	AX877037 Sequence
32	1575.5	40.7	2789	6	BD156451 Primer fo
33	1575.5	40.7	2789	9	AK001580 Homo sapi
34	1510	39.0	2801	9	BC005029 Homo sapi
35	1473	38.1	2648	9	AK025976 Homo sapi
36	1473	38.1	2793	9	AJ430349 Homo sapi
37	1422	36.7	2779	10	MMU441086 Mus muscu
38	1318.5	34.1	2261	9	BC017217 Homo sapi
39	1271	32.8	2289	10	BC003726 Mus muscu
40	1247.5	32.2	2129	6	AX772981 Sequence
41	1247.5	32.2	2129	9	HSU47926 Human unkno
42	1190	30.7	2214	10	BC016431 Mus muscu
43	1179.5	30.5	1656	6	AX772984 Sequence
44	1169	30.2	3226	9	AK126766 Homo sapi
45	1108	28.6	703	6	AX868722 Sequence

ALIGNMENTS

RESULT 1

```

AF097432      2600 bp      mRNA      linear      PRI 22-NOV-2000
LOCUS      Homo sapiens GROS1-L protein mRNA, complete cds.
DEFINITION      AF097432
ACCESSION      AF097432
VERSION        AF097432.1 GI:11127637
KEYWORDS      Homo sapiens (human)
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE      1 (bases 1 to 2600)
AUTHORS      Kaul,S.C., Sugihara,T., Yoshida,A., Nomura,H. and Wadhwa,R.
TITLE        GROS1, a potential growth suppressor on chromosome 1: its identity
to basement membrane-associated proteoglycan, leprecan
JOURNAL      Oncogene 19 (32), 3576-3583 (2000)
MEDLINE      20406927
PUBMED      10951563
REFERENCE      2 (bases 1 to 2600)
AUTHORS      Wadhwa,R., Sugihara,T. and Kaul,S.
TITLE        Direct Submission
JOURNAL      Submitted (06-OCT-1998) Group 4, Chugai Research Institute of
Molecular medicine, 153-2 Nagai, Nihari, Ibaraki 300-41, Japan
FEATURES      Location/Qualifiers
source      1..2600
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /chromosome="1"
            /tissue type="testis"
            52..2262
            /note="member of protein B family"
            /codon_start=1
            /product="GROS1-L protein"
            /protein_id="AAG31019.1"
            /db_xref="GI:11127638"
            /translation="MAVRALKLTLLAVVAASQAVESEAGKGMVTPDLLAEFTA
            AYARGDPGVLSNERALRSALRALRLRCRTQCAADFPWELDPWSPPQAQASAG
            ALRDLSPFGLLSERLRCGLRGLPAAHSLESELEFRKRSYNYLVQVAFKINKLE
            KAVAAAHFTFVNGNHEMQLNDYYQTMGSKVEADFKOLETPHMQEFLGRVLYSE
            EQPEAVPHLEALQEVFAVECRALCEGPDYDGYNYLEYNADLFQAITDHYIOVL
            NKNQCVTELASHPSREKPEDFLPSHYNYLQFAYNYIGNYTOAGCAKTYLLFPND
            EVMQNLYAYAAAGBEHTRSIGPRESAKVRSRSLLEKELFFAYDVFQIPVDPS
            WTPPEVTPKLOEKSKERSTAVRISOEIGNLMKELETVEEKTESLDVSRLTRSG
            PLLYEGISLTWNKSLNGYRVWMDGVSDHECQELQRLNNVAATSGDVRGTSPT
            PNEKPYGVTFKALKGOEGKVPQSAHLYNNVTEKVRRIIMESYFRDLTPLYFSYHL
            VCRPAIEVQAERKDDSHFVNDVNCILNAETLVCKPEPAYTRFDYSAILYNGDPDG
            GNFTFELDAKVTAEVPOQCGRAVGSSTENPHGVKAVTRGRCALIAFWFLDPRH
            SERDRVQADDLVKMLFSEPMWDLSEQPLDAQQGPPEPAQESLSGESKPKDEL"
ORIGIN
Alignment Scores:
Pred. No.:      2,44e-295      Length:      2600
Score:          3870.00      Matches:      736
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     100.00%      Indels:      0
DB:              9      Gaps:      0
US-10-045-815-4 (1-736) x AF097432 (1-2600)
Qy      1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20
Db      52 ATGGCGGTACGCGGTGAAGCTCTGACACACTGCTGGCTGCTGGCGGCTGCTCC 111
Qy      21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40
Db      112 CAAGCCGAGGTGAGTCCGAGGAGGATGGGATGGTGGATGGCTGCTTCCTCGCC 171
Qy      41 GluGlyThrAlaAlaTrpAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60
Db      172 GAGGGGACCGCAGCTACGCGCGCGGGAGCTGGCCGCGGTGGTCTGAGCATGGAACG 231
Qy      61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80

```

```

232 GGCTCGCTCCCGGAGCCCTCCGCGCCCTTCGCTCGCTCGCTCGGACCCAGGTGTGCC 291
Qy      81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100
292 GCGGACTTCCGCTGGAGCTGACCCCGACTGGTCCCCCAGCCCGCCAGGCTCGGGC 351
Qy      101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120
352 GCGGCGCCCTCGGAGACCTGAGCTTCTTCGGGGGCGCTTCGCGCGCTGCGCTGCTG 411
Qy      121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe 140
412 CGCGCTGCTCGGCGCGCGCGCCACTCGCTCAGGAAGAGATGGAGCTGGAGTTC 471
Qy      141 ArgLysArgSerProTrpAsnTrpLeuGlnValAlaTrpPheLysIleAsnLysLeuGlu 160
472 CGCAAGCGGAGCCCTACAACTACCTGAGGTCGCTACTTCAAGATCAACAAGTTGAG 531
Qy      161 LysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetGln 180
532 AAGAGCTGTGCTGCAGCACACACCTTCTTCGTGGCAATCTCTGAGCACATGGAATGAG 591
Qy      181 GlnAsnLeuAspTrpTrpGlnThrMetSerGlyValLysGluAlaAspPheLysLeu 200
592 CAGAACCTAGACTATTACCAACCATGCTGGAGTGAAGAGGCGGCACTTCAAGGATCTT 651
Qy      201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTrpSerGluGluGln 220
652 GAGACTCAACCCCATATGCAAGATTTGCACTGGGAGTGGCACTTACTTCAGGAAGACAG 711
Qy      221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTrpPheValAlaTrpGlu 240
712 CCACAGGAAGCTGTGCCCCACCTAGAGCGGCGCTGCAAGAATACTTTGTGGCTATGAG 771
Qy      241 GluCysArgAlaLeuCysGluGlyProTrpAspTrpAspGlyTrpAsnTrpLeuGluTrp 260
772 GAGTCCGCTGCGCTCTGGAAGGGCGCTTACTGACTACGATGCTCAACTACTCTTGAGTAC 831
Qy      261 AsnAlaAspLeuPheGlnAlaAlaThrAspHisTrpIleGlnValLeuAsnCysLysGln 280
832 AACGCTGACCTCTCCAGGCCATCACACATCATATCATCCAGGTCTCTCAACTGTGAAGCAG 891
Qy      281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300
892 AACTGTGTCCGAGCTTGTCTCCACCACCAAGTCGAGAAGAGCCCTTTGAAGACTTCTCTC 951
Qy      301 ProSerHisTrpAsnTrpLeuGlnPheAlaTrpTrpAsnIleGlyAsnTrpThrGlnAla 320
952 CCATCGCATTAATATCTGAGTTTGCCTACTATAACATTTGGGAATTTATACACAGCT 1011
Qy      321 GlyGluCysAlaLysThrTrpLeuLeuPhePheProLeuAspGluValMetAsnGlnAsn 340
1012 GGTGAATGTGCCAGACCATCTCTCTCTCTCTCCCAATGACGAGGTGATGAACCAAAAT 1071
Qy      341 LeuAlaTrpTrpAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360
1072 TTGGCTATTATGAGCTATGCTTTGGAGAAGAACACACACAGATCTCATCGGCCCGCTGAG 1131
Qy      361 SerAlaLysGluTrpArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTrp 380
1132 AGTCCAAAGGAGTACCGACAGCAAGCTACTGCGAAAAGAACTGCTTTTCTTCGCTTAT 1191
Qy      381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400
1192 GATGTTTTTGAATTTCCCTTTGTGATCCGGATTTCATGGACTCCAGAGAAGATGATTCCTC 1251
Qy      401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420
1252 AAGATTGCAAGAGAAACAGAAAGTCAGAACCGGAAACAGCCGTACGCATCTCCAGGAG 1311
Qy      421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440

```

Db 1312 ATTGGGAACCTTATGAGGAATTCGAGACCCCTTGTGGAGAGAACCAAGAGTCACTG 1371  
Qy 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyLeuSerLeuThr 460  
Db 1372 GATGTGACGACATGACCCGGGAAGTGGCCCTGCTGTATGAAGGCATCACTCTCAC 1431  
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValMetAspGlyValIleSerAsp 480  
Db 1432 ATGNACTCCAACTCTCAATGTTTACACGCGGTGTGTATGACGGGTAACTCTTGAC 1491  
Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
Db 1492 CACGAGTGTGAGGAGTGCAGAGACTGACCAATGTGGCAACCTCAGGAGATGGCTAC 1551  
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
Db 1552 CGGGGTGAGACCTCCACATACCTCCCAATGAAAGTTCATGGTGTCACTGCTTCAAA 1611  
Qy 521 AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
Db 1612 GCCCTCAAGCTGGGGCAAGGCAAGTTCCTCTGCGAGAGTGCCACCTGTACTCAAC 1671  
Qy 541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
Db 1672 GTGACGAGAGAAAGTGGCGGCATCATGGAGTCTCTACTTCGCGCTGGATACGCCCTCTAC 1731  
Qy 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArgLys 580  
Db 1732 TTTTCTTACTCTATCTGTGTGCGGACCTGCCATCGAAGGTCCAGGACAGAGGAAG 1791  
Qy 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
Db 1792 GATGATAGTATCCAGTCCACGTGACCACTGATGCTGTAATGCCGAGACCTCGTGTGT 1851  
Qy 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
Db 1852 GTCAAGAGCCCCAGGCTTACACCTTCGCGACTACAGCGCCATCTTTTACCTTAATGGG 1911  
Qy 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
Db 1912 GACTTCGATGCGGAAACTTTTATTTCACTGAACCTGATGCCAAGACCGTGACGGCAGAG 1971  
Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
Db 1972 GTGCAGGCTGTGAGGAGCGCGTGGGATTCCTTCAGGCACTGAAACCCACATGGA 2031  
Qy 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680  
Db 2032 GTGAGGCTGTACACGGGGGCGAGCGCTGTGCGCATCGCCCTGTGGTTTCACTTGGACCT 2091  
Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
Db 2092 CGACACAGCGAGCGGAGCAGGCTGCAGGCAGATGACCTGGTGAAGATGCTCTTCAGCCCA 2151  
Qy 701 GluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
Db 2152 GAAGAGATGGACCTCTCCAGAGACAGCCCTGATGCCAGAGCGGCGCCCGGCAACT 2211  
Qy 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2212 GCACAAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGATGAGCTA 2259

RESULT 2  
AX136191  
LOCUS AX136191 2563 bp DNA linear PAT 30-MAY-2001  
DEFINITION Sequence 113 from Patent EP1067182.  
ACCESSION AX136191  
VERSION AX136191.1 GI:14272599  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source

CDS

ORIGIN

Alignment Scores:  
Pred. No.: 1,87e-293 Length: 2563  
Score: 3846.00 Matches: 733  
Percent Similarity: 99.59% Conservative: 0  
Best Local Similarity: 99.59% Mismatches: 3  
Query Match: 99.38% Indels: 0  
DB: 6 Gaps: 0  
US-10-045-815-4 (1-736) x AX136191 (1-2563)

Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaAlaSer 20  
Db 31 ATGGCGGTACGCGGCTTGAAGTGTGACACACATGCTGGCTGTGCTGGCGCTGCCTCC 90  
Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla 40  
Db 91 CAAGCGAGTCCAGTCCGAGGAGGATGGGATGCTGACGCTGATCTCTTCGCC 150  
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60  
Db 151 GAGGGACCGACGCTTACGCGCGGAGTGGCGGGGCTGCTGAGCATGGAAACGG 210  
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
Db 211 GCGCTGCGCTCCGCGGACGCTCCGCGGCTTCCGCTGCGCTGCGGACCCAGTGTGCC 270  
Qy 81 AlaAspPheProTyrProTyrLeuAspProAspTyrProSerProSerProAlaGlnAlaSerGly 100  
Db 271 GCGGACCTTCCGTTGGAGCTGACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 330  
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLeu 120  
Db 331 GCGCGCGCTGCGGACCTGAGCTTCTTCGGGGGCTTCTGCTGCTGCTGCTGCTGCTGCTG 390  
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe 140  
Db 391 CGCGCTGCTGCGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 450  
Qy 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
Db 451 CGCAAGCGGACCCCTTACAACTTCTGAGGTGCGCTTCTTCAAGATCAACAGTTGGAG 510

1  
Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and  
Hayashi,K.  
Secretory protein or membrane protein  
Patent: EP 1067182-A 113 10-JAN-2001;  
Helix Research Institute (JP)  
Location/Qualifiers  
1..2563  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
31..2241  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC39738.1"  
/db\_xref="GI:14272600"  
/db\_xref="REMBEL:CAC39738"  
/translation="MAVRALKLITLLAVAAASQAQVESEAGNMTPTDILLPAGETA  
AYARDGFGVLSMERALRSALRLRQTCADFPWELDPDMSFPAQSGAA  
ALRDSFFGGLLRRAACRLCPYPAHSLSEMELEFRKESPYNQVAYFKNKLE  
KAVAAAHFTFFGNPEHMEMQNLDDYQTMGSGKADFDKLETPHQMRFLGRLYSE  
POPOAVPHLEAALQYFVAYEECRALCEGPDYDGYNLEYNADLFOAITDHYIQVL  
NKONCVTELAHSPREKPFEDFLPHYNYLQFAYVNIQVNTQAVECAKTYLLPFEND  
EVMNQNLAYAAALGSEHSTRSIGPRESAKEYRQSLLEKELLFFAYDVGFPVDPDS  
WTPPEVIRLQEKOKRSERATVRISQISGNLMEIEFTLVEEKIKESLDVSRLLREGG  
PLLYEGISLTMNSKLLNGSRVMDGVISDHECOELQRLTNVAATSGGYRGYQSPHT  
PNEKEYVTVFALKLQEGKVPQSAHLYYNVTKEKRRIMESFRLLPTPIFYSHL  
VCRTAIEVQAEKDDSHPVNDICLNALTYLVCVKEPPAYTFRDYSAILYNGFDG  
GNFYPTLDAKTVAEVOQCGRAVGFSSGTENPHGVKAVTRGQRCALALWFTLDPRH  
SERDRVQADDLVKMLFSPEMDLSQEQPLDAQGPPEPAQESLSSESSEKXDEL"

QY 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180  
 DB 511 AAAGCTGTGTGTCAGACACACACCTTCTCGTGGGCAATCTCGAGCACATGAAATGCAG 570  
 QY 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
 DB 571 CAGAACCTAGACTATTACCAACCACTGCTGGAGTGAAGAGGCGGACTTCAAGGATCTT 630  
 QY 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValAlaArgLeuTyrSerGluGlnGln 220  
 DB 631 GAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTGCAGCTCTACTCAGAGAGAAACAG 690  
 QY 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
 DB 691 CCACAGGAGAGCTGTGCCCACTAGAGCGCGCTGCAAGAAATACTTTGGCCCTATGAG 750  
 QY 241 GluCysArgAlaLeuGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
 DB 751 GAGTCCCGTGCCTCTGCGAAGGCGCTTATGACTACGATGCTCAACACTACCTTGAATAC 810  
 QY 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
 DB 811 AACGCTGACCTCTTCCAGGCCATCAGATCATATACATCCAGGTCCTCAACTGTAAAGCAG 870  
 QY 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
 DB 871 AACTGTGTCCAGGAGCTTGTCCCAAGTTCAGAGAGAGCCCTTTGAAGACTTCTCCTC 930  
 QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
 DB 931 CCATCGCATTAATAATATCTGAGTTTGCCTACTATAACATTGGGAATATATACACAGGCT 990  
 QY 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
 DB 991 GTTGAAATGTGCAAGACCTATCTCTTCTTCCCAATGACGAGGTGATGAACCAAAAT 1050  
 QY 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
 DB 1051 TTGGCTATTATGACGTATGCTTGGAGAAACACACACATCCATCGGCCCGCTGAG 1110  
 QY 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr 380  
 DB 1111 AGTGCAGAGGATACCGACGAGAGCTACTGGAAAGAACTGCTTTTCTTCGCTTAT 1170  
 QY 381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400  
 DB 1171 GATGTTTTTGGAAATTCCTTTTGTGATCCGATTCATGGACTCCAGAGAAAGTATTCCTC 1230  
 QY 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
 DB 1231 AAGAGATTGCAAGAGAAACAGAAAGTCAGAACGGGAAACAGCCGATCCATCTCCAGGAG 1290  
 QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
 DB 1291 ATTGGGAACCTTATGAAGAAATCGAGACCCCTTGTGGAAGAGAGACCAAGGAGTCACTG 1350  
 QY 441 AspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
 DB 1351 GATGTGACAGACTGACCCGGAGAGTGGCCCTCTGTATGAAGGCATCAGTCTCAAC 1410  
 QY 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
 DB 1411 ATGAATCCAACTCTGAAATGGTTCCAGCGGTGTGATGAGCGGCTAATCTCTGAC 1470  
 QY 481 HisGluCysGlnGluLeuLeuArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
 DB 1471 CACGAGTGTACGAGGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGCTAC 1530  
 QY 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
 DB 1531 CGGGGTGAGACTTCCCACTATCTCCCAATGAAGAAAGTTCTATGTTGATCTGTCTCAA 1590

QY 521 AlaLeuLysLeuGlyGlnGluGlyValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
 DB 1591 GCCTCAAGCTGGGCAAGAGGCAAAAGTTCTCTGCAAGTGCACCTGTACTACAC 1650  
 QY 541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
 DB 1651 GTGACGAGAGGTGGCGGCATCATGAGTCTCTACTTCCGCTCGATACGCCCTCTAC 1710  
 QY 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
 DB 1711 TTTTCTACTCTCATCTGTGTGGCGCACTGCGCATCGAAGAGTCCAGCAGAGAGGAG 1770  
 QY 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
 DB 1771 GATGATAGTCATCCAGTCCACGTGGACCACTGATCTCTGAATGCCGAGACCTCTGTGT 1830  
 QY 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
 DB 1831 GTCAAGAGCCGCCACCTACACCTTCCGCACTACAGCGCATCTTACCTAAATGGG 1890  
 QY 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
 DB 1891 GACTTCGATGGCGGAACCTTTTATTTCTGAACTGGATGCCAAGACCGTGACGCGCAG 1950  
 QY 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
 DB 1951 GTGACGCTCAGTGTGGAAGAGCCGTGGATTTCTTCAAGGCACTGAAACCCACATGA 2010  
 QY 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680  
 DB 2011 GTGAAGGCTGTCAACAGGGGCGAGCGCTGTGCACTGGCCCTGTGTTTCCCTTGACCT 2070  
 QY 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
 DB 2071 CGACACAGCGAGCGGACAGGGTGCAGCAGATGATGCTGGTGAAGATGCTCTTCAGCCCA 2130  
 QY 701 GluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
 DB 2131 GAAGAGATGAGCTCTCCAGAGAGAGCCCTCGGATGCGCAGGGGCCCCCGAACCT 2190  
 QY 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
 DB 2191 GCACAGAGCTCTCTCTCAGGAGTGAATCGAAGCCCAAGATGAGCTA 2238  
 RESULT 3  
 BD123548  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 BD123548  
 Secretory protein or membrane protein.  
 BD123548  
 BD123548.1 GI:23218493  
 JP 2002017376-A/57.  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and  
 Hayashi,K.  
 1 (bases 1 to 2563)  
 Secretory protein or membrane protein  
 Patent: JP 2002017376-A 57 22-JAN-2002;  
 HELIX RESEARCH INSTITUTE  
 OS Homo sapiens (human)  
 PN JP 2002017376-A/57  
 PD 22-JAN-2002  
 PF 07-JUL-2000 JP 2000253173  
 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, YURI KAWAI, TOMOYASU  
 SUGIYAMA,  
 PI KOJI HAYASHI  
 PC  
 C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC  
 10, C12P21/02, C12Q1/68//C12P21/08, C12N15/00, C12N5/00 CC  
 Secretory protein or membrane protein



[illegible]



811 AACGCTGACCTCTTCAGGCGCATCAGAGATCATTACATCAGGCTCCTCAACTGTGAAGCAG 870  
Qy AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
Db AACTGTGTACGAGCTTGTCTCCACCAAGTCGAGAGAGCCCTTTTGAAGACTTCCTC 930  
Qy ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
Db CCATCGCATTAATAATTATCTCGAGTTTGCTTACTATAACATTGGGAATTATACACAGGCT 990  
Qy 321 GLYGLuCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
Db 991 GTTGAATGTCGAAGACCTATCTTCTCTCTCTCCCAATGACAGAGTGATGAACCAAT 1050  
Qy 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
Db 1051 TTGGCCCTATTATGAGCTATGCTTGAGAGAAACACACACAGATCCATCGCCGCCCTGAG 1110  
Qy 361 SerAlaIysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhePheAlaTyr 380  
Db 1111 AGTGCCAAGAGTACCGACAGCGAAGCTTACTCGAAAGAACTGCTTTCTTCGCTTAT 1170  
Qy 381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400  
Db 1171 GATGTTTTCGAATTCCTTTGTGGATTCGGAATTCGACTCCAGAGAGATTCCTCC 1230  
Qy 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
Db 1231 AAGAGATTGCAAGAGAAACAGAGTCCAGACGGGAACACCGCTACGCTATCCCGAGGAG 1290  
Qy 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluLysThrLysGluSerLeu 440  
Db 1291 ATTGGGACCTTATGAGGAATTCGAGACCTTGTGGAGAGAGACCAAGAGTCACTG 1350  
Qy 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
Db 1351 GATGTGAGCAGACTGACCCGGAAGTGGCCCTGCTGTATGAAGGCATCACTTCACC 1410  
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
Db 1411 ATGAACCTCAAACTCTCGAATGTTTCCAGCGGTGTGTGAGCGGCTATCTCTGAC 1470  
Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
Db 1471 CACGAGTGTGAGGAGCTGCAGAGACTGACCAATGTGGCAACCTCAGGAGATGGCTAC 1530  
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
Db 1531 CGGGGTGAGACCTCCCCACATACTCCCAATGAAAGTTCTATGGTGTCACTGTCTTCAA 1590  
Qy 521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
Db 1591 GCCCTCAAGCTGGGGCAGAGGCAAGATTCCTCTGCAGAGTGCACACCTGTACTACAC 1650  
Qy 541 ValThrGluLysValArgAlaGlyMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
Db 1651 GTGACGAGAGAGTGGCGCGCATCATGGAGTCTTACTTCGCGCTGGATACGCCCTCTAC 1710  
Qy 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
Db 1711 TTTTTCCTACTCTCATCTGGTGTGGCGGACCTGCCATCGAAGAGTCCAGCGAGAGGAG 1770  
Qy 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
Db 1771 GATCATAGTCCATCCAGTCCAGTGGACAACATGTCATCTGAATGCCAGAGACCTCGTGT 1830  
Qy 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
Db 1831 GTCAAG 1890  
Qy 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
Db 1891 GACTTCGATGCGGGAACCTTTTATTTTCACTGAAGTGGATGCCAAGACCGTGACGGCAGAG 1950

Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
Db 1951 GTGACAGCCTCAGTGTGGAAGAGCCGTGGGATTTCTTTTTCAGGCACTGAAACCCCATGGA 2010  
Qy 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680  
Db 2011 GTGAAGGCTGTACACAGGGGCGAGCGCTGTGCCATCGCCCTGTGGTTCACCTGACCT 2070  
Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
Db 2071 CGACACAGCGAGCGGACAGCGGTGCAGGCAGATGACCTGGTGAAGATGCTCTTCAGCCCA 2130  
Qy 701 GluGluMetAspLeuSerGlnGluAlaProLeuAspAlaGlnGlnGlyProGluPro 720  
Db 2131 GAAGAGATGGACCTTCTCCAGAGAGCGCCCTGTGATGCCAGAGGGGCCCCCGAACCT 2190  
Qy 721 AlaGlnGluSerLeuSerGlySerGlySerLysProLysAspGluLeu 736  
Db 2191 GCACAGAGATGCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2238

## RESULT 5

AX704765 2583 bp DNA linear PAT 04-APR-2003  
LOCUS Sequence 69 from Patent WO0202634.

## DEFINITION

AX704765

## ACCESSION

AX704765.1 GI:29561431

## VERSION

Homo sapiens (human)

## KEYWORDS

Homo sapiens

## ORGANISM

Homo sapiens

## REFERENCE

1

Gururajan, R., Hafalia, A.J., Kallick, D.A., Patterson, C., Azimzai, Y.,

Khan, F.A., Xu, Y., Yao, M.G., Yue, H., Au-Young, J., Batra, S., Y.T.,

Baughn, M.R., Borowsky, E.A., Lo, I.F., Lu, D.A., Lu, Y., Tang, Y.T.,

Yang, J., Zingler, K.A., Deleane, A.M., Gietzen, K., Marcus, G.A.,

Nguyen, D.B., Policky, J.L., Ramkumar, J., Thangavelu, K., Walla, N.K.

and Warren, B.A.

Human extracellular matrix and cell adhesion polypeptides

Patent: WO 0202634-A 69 10-JAN-2002;

INCITE GENOMICS INC. (US)

## FEATURES

Location/Qualifiers

1..2583

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

/note="incyte ID No: 6780147CB1"

## ORIGIN

Alignment Scores:

Pred. No.: 1

Score: 1.89e-293

Length: 2583

Matches: 3946.00

Conservative: 0

Percent Similarity: 99.59%

Mismatches: 3

Best Local Similarity: 99.58%

Indels: 0

Query Match: 99.38%

Gaps: 6

US-10-045-815-4 (1-736) x AX704765 (1-2583)

Qy 1 MetAlaValArgAlaLeuLysLeuThrThrLeuAlaValAlaAlaSer 20

Db 42 ATGGCGGTACGCGGTGTGAAGTCTGACCACTGCTGGCTGTCTGGCCCTGCTCC 101

Qy 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40

Db 102 CAAGCCGAGTTCGAGTCCGAGGAGGATGGGCGATGTTGACGCTGATCTCTTCGCC 161

Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60

Db 162 GAGGGGACCGCAGCTTACCGCGCGGAGTACCGCGGGGCTGGCCCGGGTGGTCTGAGCATGGAACGG 221

Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80

Db 222 GGGCTGCGCTCCGGGAGCCCTCGCGCCCTTCGCTGCGCTGCGGACCCAGTGTGCC 281  
Qy 81 AlaAspPheProTyrGluLeuAspProAspTyrPheSerProAlaGlnAlaSerGly 100  
Db 282 GCGGACTTCCCGTGGAGCTGAGCCCGGACTGGTCCCGGAGGCGGCGGCGGCTCGGGC 341  
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLeu 120  
Db 342 GCGCGCGCCCTCGCGGACTGAGCTTCTCGGGGCGCTTCGCGTGGCTGCGCTG 401  
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
Db 402 GCGCGCTGCTGGGCGCGCGCGCCGCGCTCGCTGAGGAGAGATGGAGCTGGAGTTC 461  
Qy 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
Db 462 CGCAAGCGAGCCCTCAACTACCTGAGGTGCGCTGCTTCAAGATCAACAGTTGGAG 521  
Qy 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180  
Db 522 AAAGCTGTGTGTCAGACACACACTTCTTCGTGGGCAATCTGAGCACATGGAATGCAG 581  
Qy 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
Db 582 CAGAACCTAGACTATTACCAACCATGCTCGAGTGAAGGAGCGGACTTCAAGGATCTT 641  
Qy 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220  
Db 642 GAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTGGGCTACTCTACTCAGAGCAACAG 701  
Qy 221 ProGlnGluAlaValProHisGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
Db 702 CCACAGGAAGCTGTGCGCCCTCAGAGCGCGCTGCAAGAAATACTTTGTGGCCATGAG 761  
Qy 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
Db 762 GAGTGGCGTGGCTCTCGGAAGGCGCTTATGACTACGATGCGTACAACTACCTTGAGTAC 821  
Qy 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
Db 822 AACGCTGACCTCTTCCAGGCGCATCACAGATCATTACATCCAGGTGCTCAACTGTAAAGCAG 881  
Qy 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
Db 882 AACTGTGTCAAGGAGCTGTCTCCCAAGTCCAGAGGAGCGCTTTGAGACTTCTCTC 941  
Qy 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
Db 942 CCATCGCATTAATAATTATCTGAGTTTGCCTTACTATAACATTGGGAATTATACACAGGCT 1001  
Qy 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
Db 1002 GTTGAAATGTCAAGACTATCTTCTCTTCTTCCCAATGACGAGGTGATGAACCAAAAT 1061  
Qy 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
Db 1062 TTGGCTTATTATGACGATATGCTTGGAGAGAAACACACACAGATCCATCGGCGCCCGTGG 1121  
Qy 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr 380  
Db 1122 AGTCCCAAGGAGTACCCAGCAGGAGGCTTACTGGAAGAAAGAACTGCTTTTCTCGCTTAT 1181  
Qy 381 AspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIlePro 400  
Db 1182 GATGTTTTTGGAAATTCCTTTGTGATCCGGATTCATGGACTCCAGAGAAGTATTCCTCC 1241  
Qy 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
Db 1242 AAGAGATTGCAAGAGAAACAGAGTCAAGCGGAAACAGCCGTACGATCTCCCGAGGAG 1301  
Qy 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluLysThrLysGluSerLeu 440

Db 1302 ATTGGGAACCTTATGAAGGAAATCGAGACCCCTGTGTGGAAGAGAACCAAGAGTCACTG 1361  
Qy 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
Db 1362 GATGTAGCAGACTGACCCGGAAGTGGCCCTCTGTGTATGAAGGATCATGCTCTCACC 1421  
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValMetAspGlyValIleSerAsp 480  
Db 1422 ATGAATCTCCAACTCTGAATGGTTCCTCCAGCGGGTGGTATGAGCGGCTAATCTCTGAC 1481  
Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
Db 1482 CACGAGTGTGAGAGCTGCAGAGACTCCCAATGTGGCAGCAACCTCAGGAGATGGCTAC 1541  
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
Db 1542 CGGGGTGACACCTCCCACTACTCCCAATGAAAGTTCATGGTGTCTCTCTCTCAA 1601  
Qy 521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
Db 1602 GGCCTCAAGCTGGGGCAGAGAGCAAGTCTCTCTGAGAGTGGCCACCTGTACTACAAC 1661  
Qy 541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
Db 1662 GTGACGGAAGGTGGGCGCATCATGAGTCTTACTTCCGCTGGATACGGCCCTCTAC 1721  
Qy 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
Db 1722 TTTTCTACTCTCATCTGTGTGCGGCTGCTGCAATGAAAGAGTCCAGGCAAGAGGAG 1781  
Qy 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
Db 1782 GATGATAGTATCCAGTCCAGTCCAGTGCACATGCTCATCTGAAATGCCAGACCTCTGTGT 1841  
Qy 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
Db 1842 GTCAAGAGCCCCAGCTTACACCTTCCGCGACTACACGCGCATCTTTACTTAATGGG 1901  
Qy 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
Db 1902 GACTTCGATGGCGAAACTTTTATTCTACTGAATGGATGCCAAGACCGTGACGCGAG 1961  
Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
Db 1962 GTGACGCTCAGTGTGGAAGAGCGCTGGGATTCCTTCAGGCACTGAAACCCACATGGA 2021  
Qy 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680  
Db 2022 GTGAAGGCTGTCCAGGCGGCGAGCGCTGTGCCATCGCCCTGTGGTTCCACCTGACCCCT 2081  
Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
Db 2082 CGACACGAGCGGAGCAGGCTGAGGATGACGATGACCTGTGTGAAGATGCTCTTCAGGCCA 2141  
Qy 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
Db 2142 GAAGAGATGGACCTCTCCAGAGCAGCGCCCTGGATGCCAGAGGCGGCCCCCGAACCT 2201  
Qy 721 AlaGlnGluSerLeuSerGlySerCysSerLysProLysAspGluLeu 736  
Db 2202 GCACAAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2249

RESULT 6  
AX877288  
LOCUS  
Sequence 12193 from Patent EP1074617.  
DEFINITION  
AX877288  
ACCESSION  
AX877288.1 GI:40032024  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE	1	Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J., Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
AUTHORS		Primers for synthesizing full-length cDNA and their use
TITLE		Patent: EP 1074617-A 12193 07-FEB-2001.
JOURNAL		Research Association for Biotechnology (JP)
FEATURES		Location/Qualifiers
source	1..2585	/organism="Homo sapiens"
		/mol_type="unassigned DNA"
		/db_xref="taxon:9606"
CDS	53..2263	/note="unnamed protein product"
		/codon_start=1
		/protein_id="CAE89658.1"
		/db_xref="GI:40032025"
		/translation="MAVRALKLTLLAVVAASQAEVESEAGMGWMTVDLLFAEGTA AVARGDPVGLWSEALRSALRLRTRCQCAADFPWELDPDWSPPAASGAA ALRDLSPFGLLRACLRCLGPPAHLSSEMELEFRKSPYNLYOVAYPKINKLE KAVAAHTFFVGNPEHMEMQNDLYQTMSGVKEADFQKLETPHMQERFLGVRLYSE EQEAVPHLEALQEVYAYEBCRALCEGYDYGNYLEYNAIDFQALTDHYIQVL NCKQNCVTELASHPSEKPFEDLPSPHYNLYQFAYNIGNYTOAVECAKTYLLFFPD EVANONLAYAAMGEBHTRSIGPRESAKERYRSLLEKLLFFAYDFGIPFVDPDS WTPEVTPKLOEKOKERETAVRISOEIGNLKEIETLVEEKTKESLDVSRLTRGG PLLYEGISLTWNSKLLNGSORVMVDGVISDCEQLRLTNVAATSGDVGRTGSPHT PNKPFYGVTPFKALQCEGKVPLOSALHYNYNTEKVRIMESYFLDTPLYPSYHL VCTAIEBQAEKRDSDHPVHVDNCLINASTLVCVKEPFAITFRDISALYLNGDFDG GNFYFTEDAKTVAEYPOQGRAVGSSTENPHGVKAVTRGRCALIALWFLTDPRH SERDRVQADLVKMLFPEEMDLSEQFLDAQQPPPEPAQESLSGESPKDEL"
ORIGIN		
Alignment Scores:	1.89e-293	Length: 2585
Pred. No.:	3846.00	Matches: 733
Score:	99.59%	Conservative: 0
Percent Similarity:	99.59%	Mismatches: 3
Best Local Similarity:	99.38%	Indels: 0
Query Match:	6	Gaps: 0
DB:		
US-10-045-815-4 (1-736) x AX877288 (1-2585)		
Qy	1	MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20
Db	53	ATGGCGGTACGCGCGTTGAAGCTGCTGACACACACTGCTGGCTGTGTCGGCGCTGCGCTCC 112
Qy	21	GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40
Db	113	CAAGCCGAGGTGCGAGTCGAGCGAGGATGGGGCATGGTGCCTGATCTGCTCTTCGCC 172
Qy	41	GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60
Db	173	GAGGGGACCGCAGCCTACGCGCGGGGACTGGCCCGGGGTGTCCTGAGCATGGAAACGG 232
Qy	61	AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
Db	233	CGCTGCGCTCCCGGACGCGCTCCGCGCCCTCGCGCTGCGCTGCGCGCCACCGATGTGCC 292
Qy	81	AlaAspPheProTyrGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100
Db	293	CGCGACTTCCCGTGGAGCTGGAACCCGACTGTGTCGCCCGCCGCGCGCGCTCGGGC 352
Qy	101	AlaGlyAlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120
Db	353	CGCGCGCGCTCGCGGACCTGAGCTTCTTCGGGGCGCTTCTGCGCTGCGCTGCGCTG 412
Qy	121	ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140
Db	413	CGCGCGCTCGCGCGCGCGCGCGCGCTGCTGCTGAGCGAAGAGATGGAGCTCGAGTTC 472
Qy	141	ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160
Db	473	CGCAAGCGGAGCCCTACAACACTACTGTCAGGTGCGCTACTTCAAGATCAACAACTGGAG 532
Qy	161	LysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetGln 180
Db	533	AAAGCTGTGTGTCAGCAGCACACACTTCTTCGTGGGCAATCTCTGAGCACATGGAATGAG 592
Qy	181	GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200
Db	593	CAGAACCTAGACTATTACCAACCAATGTCTGGAGTGAAGGAGCGCGACTTCAAGATCTT 652
Qy	201	GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGln 220
Db	653	GAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTGGCACTTACTACAGAGAACAG 712
Qy	221	ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240
Db	713	CCACAGGAAGCTGTGCCCCACCTAGAGCGCGCTGCAAGAAATACCTTTGTGGCCATGAG 772
Qy	241	GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260
Db	773	GAGTGGCGTCCCTCTCGGAAGGGCCCTATGACTACGATGGCTACCACTACCTTGGATAC 832
Qy	261	AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280
Db	833	AACGCTGACCTCTCCAGGCCATCAGATCATATCATCCAGGTCTCTCAACTGTGAAGCAG 892
Qy	281	AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300
Db	893	AACTGTGTCAGGAGCTTCTCCCAACCACTCGAGAGAGCCCTTTGAAGACTTCTCTC 952
Qy	301	ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320
Db	953	CCATGCGATTATTAATATCTGCAGTTTGCCTACTATAACATTTGGGAATTTATACAGGCT 1012
Qy	321	GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340
Db	1013	GTTGAATGGCCAGACACTATCTTCTTCTTCCCATGAGGAGTGTATGAACCAAAAT 1072
Qy	341	LeuAlaTyrTyrAlaAlaMetLeuGluGluHisThrArgSerIleGlyProArgGlu 360
Db	1073	TTGGCCTATTATGACGTATGCTTGAGAAAGAACACACAGATCCATCGGCCCGCTGAG 1132
Qy	361	SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhePheAlaTyr 380
Db	1133	AGTGGCAAGAGTACCGAGCGAAGCTTCTGGAAGAAAGAACTCTTCTTCTCGCTTAT 1192
Qy	381	AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400
Db	1193	GATGTTTTTGGAAATCCCTTTTGGATCGGATTCATGGACTCCAGAGAAAGTATCC 1252
Qy	401	LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420
Db	1253	AAAGATTTCAGAGAGAAACAGAAAGTCAGAACCGGAAACAGCGCTACGCATCTCCCGAGG 1312
Qy	421	IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440
Db	1313	ATTGGGAACCTTATGAGGAAATCGAGACCTTGTGGAGAGAGAGACCCAGAGTCACTG 1372
Qy	441	AspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyIleSerLeuThr 460
Db	1373	GATGTGAGCAGACTGACCCCGGAAGTGGCCCTCTGTATGAGGCAATCATCTCTCAC 1432
Qy	461	MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480
Db	1433	ATGAATCCCAACTCTCGAATGTTCCCGCGGGTGGTATGACCGCGCTATCTCTGAC 1492
Qy	481	HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500
Db	1493	CAGGAGTGTGAGGAGCTGAGAGACTGACCAATGTGGCAAGCAACCTCAGAGATGGCTAC 1552
Qy	501	ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520
Db	1553	CGGGGTGACAGCTCCCAATCATCTCCCAATGAAAGTCTTCTGGTGTCTACTGTCTTCAA 1612
Qy	521	AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540

Db	1613	GCCCTCAAGCTGGGGCAAGAGCAAGTTCCTCTGCAGAGTGCCCACTGTACTACAAC	1672
Qy	541	ValThrGluIysValArgArgGilewMetGluSerTyrPheArgLeuAspThrProLeuTyr	560
Db	1673	GTGACGGAGAGGTGGCGGCATCATGAGTCTCTACTTCGCCCTGGATACGCCCTCTAC	1732
Qy	561	PheSerTyrSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArgLys	580
Db	1733	TTTTTCTACTCTCATCTGGTGTGGCGCACTGCCATCGAAGAGGTCTCAGGAGAGGAAG	1792
Qy	581	AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys	600
Db	1793	GATGATAGTCATCCAGTCCACCTGGACACATGCAATCTGTAATGCCGAGACCTCGTGTGT	1852
Qy	601	ValIysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly	620
Db	1853	GTCAAGAGAGCCCGAGCCTACACCTTCGCGCACTACAGCGCCATCTTTACCTAAATGGG	1912
Qy	621	AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaIysThrValThrAlaGlu	640
Db	1913	GACTTCGATGCGGAAACTTTATTTTCACTGAATCTGATGCCAAGACCGTGACGGCAGAG	1972
Qy	641	ValGlnProGlnCysGlyArgAlaValAlaGlyPheSerSerGlyThrGluAsnProHisGly	660
Db	1973	GTGCAGCCTCAGTGTGGAAGAGCGGTGGATCTCTTCAGGCACCTGAAACCCACATGGA	2032
Qy	661	ValIysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro	680
Db	2033	GTGAAGGCTGTACACAGGGGCGACGTGTGCCATCGCCCTGTGGTTCACCCCTGGACCT	2092
Qy	681	ArgHisSerGluArgAspArgValGlnAlaAspAsnLeuValIysMetLeuPheSerPro	700
Db	2093	CGACACAGCAGCGGACAGGGTCAGGCAGATGACCTGTGTGAAGATGCTCTTCAGCCCA	2152
Qy	701	GluGluMetAspLeuSerGlnGluInProLeuAspAlaGlnGlnGlyProGluPro	720
Db	2153	GAGAGATGACACTCTCCAGGAGACGCCCTCGATGCCAGAGCGCCCCCGCAACCT	2212
Qy	721	AlaGlnGluSerLeuSerGlySerGluSerLysProIysAspGluLeu	736
Db	2213	GCACAGAGTCTCTCTCAGCAGTGAATCGAAGCCCAAGATGAGCTA	2260
RESULT 7			
BD156581		2585 bp DNA linear	PAT 17-JAN-2003
LOCUS			
DEFINITION		Primer for synthesizing full-length cDNA and use thereof.	
ACCESSION		BD156581	
VERSION		BD156581.1	GI:27862339
KEYWORDS		JP 2002191363-A/11424	
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE		1 (Bases 1 to 2585)	
JOURNAL		Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J., Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T. Primer for synthesizing full-length cDNA and use thereof Patent: JP 2002191363-A 11424 09-JUL-2002; HELIX RESEARCH INSTITUTE	
COMMENT		OS Homo sapiens (human)	
		PN JP 2002191363-A/11424	
		PD 09-JUL-2002	
		PF 28-JUL-2000 JP 2000280990	
		PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU SAITO,	
		PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU, KEIICHI NAGAI,TETSUJI OTSUKI	
		PC	
		C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/10,	PC
		C12P21/02, C12P21/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC	
		Primer for synthesizing full-length cDNA and use thereof FH Key	

833 AACGCTGACCTCTTCAGGCGCATACAGATCAATTACATCCAGTCTCAACTGTAAACAG 892  
281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysPheProGluAspPheLeu 300  
893 AACTGTGTCCAGGAGTGTCTTCCCAAGTCGAGAGAGCCCTTTGAAGACATCTCTC 952  
301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
953 CCATCGCATTAATATATCTGCACTTTCCTTCTTCCCAATGACGAGGTGATGAACCAAAAT 1012  
321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
1013 GTTGAATGTCCAGACCTATCTTCTTCTTCCCAATGACGAGGTGATGAACCAAAAT 1072  
341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
1073 TTGGCCATTAATGATGCTTGTGGAGAGACACACACAGATCCATCGGCCCTCGTGG 1132  
361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaTyr 380  
1133 AGTCCCAAGAGTACCCAGACGAGGCTACTGCTGAAAAGAACTGCTTTCTTCGCTTAT 1192  
381 AspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIlePro 400  
1193 GATGTTTTTGGAAATCTTCTTGTGATCCGGAATTCATGGACTCCAGAGAGTGAATCC 1252  
401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
1253 AAGAGATTCAGAGAAACAGAGTCAAGAGCGGAAACACCGCTAGCATCTCCAGGAG 1312  
421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluLysThrLysGluSerLeu 440  
1313 ATTGGGAACCTTATGAAGGAAATCGAGACCTTGTGGAGAGAGACCAAGAGTCACTG 1372  
441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
1373 GATGTGAGCAGACTGACCCGGAGGTGGCCCTGCTGTATGAGGCATCACTCTCAC 1432  
461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
1433 ATGAACCTCAAACTCTCTGAATGTTTCCAGCGGTGTGTGATGACGCGCTAATCTCTGAC 1492  
481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
1493 CACAGTGTGAGGAGTGGCGGCGATCATGGAGTCTTCTTCCGCTGGATACGCCCTCTAC 1552  
501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
1553 CGGGGTGAGACTCCCGACATCTCTCCCAATGAAAGTTCTATGTGTCACTGTCTTCAA 1612  
521 AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
1613 GCCCTCAAGCTGGGGCAAGAGGCAAGTTTCTCTGAGAGTGCCTGCTGAGTCACTCAAC 1672  
541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
1673 GTGACGAGAGAGTGGCGGCGATCATGGAGTCTTCTTCCGCTGGATACGCCCTCTAC 1732  
561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
1733 TTTTCTCTACTCTCATCTGCTGCTGCGCACTGCCATCGAAGAGTCTCCAGGAGAGGAG 1792  
581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
1793 GATGATAGTATCCAGTCCAGTCCAGCAACTGCAATCTGATGCGGAGACCTCTGCTGT 1852  
601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
1853 GTCAAGAGGCCCCAGCGCTACACCTTCCGCGACTACAGCGCATCTCTTACCTAAATGGG 1912  
621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
1913 GACTTCGATGCGGAAACTTTTATTTCACTGAATGCTGCAAGCGTCAAGCGGAGAG 1972

641 ValGlnProGlnCysGlyValArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
1973 GTGACGCTCTAGTGTGGAGAGCGCTGGGATCTCTTCAGGACCTGAAACCCCATGGA 2032  
661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680  
2033 GTGAAGGCTGTCCACAGGGGCGAGCGCTGTGCCATCGCCTGTGTTTCCCTGACCT 2092  
681 ArgHisSerGluArgAspValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
2093 CCACACACGAGCGGAGAGGTCAGGTCAGGACAGATGACCTGTGTGAGATGCTCTTCAGCCCA 2152  
701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
2153 GAAGAGATGACCTCTCCAGGAGCAGCGCTGTGATGCCAGCAGCGGCCCCCGAACCT 2212  
721 AlaGlnGluSerLeuSerGlySerLeuSerLysProLysAspGluLeu 736  
2213 GCACAAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2260

## RESULT 8

AK027697 2585 bp mRNA linear PRI 01-AUG-2002  
LOCUS Homo sapiens cDNA FLJ14791 fis, clone NT2RP4001064, weakly similar  
DEFINITION to SYNAPTONEMAL COMPLEX PROTEIN SC65.

AK027697  
ACCESSION AK027697.1 GI:14042570

VERSION oligo capping; fis (full insert sequence).

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

## AUTHORS

Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,  
Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y.,  
Matsumura,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,  
Wakamatsu,A., Nakamura,Y., Nagahara,K., Masuho,Y. and Sasaki,N.  
NEDO human cDNA sequencing project

Unpublished

2 (bases 1 to 2585)

Isogai,T. and Otsuki,T.

Direct Submission

Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail:genom.cs@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)

NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing,  
Research Association for Biotechnology; cDNA library construction,  
5'- & 3'-end one pass sequencing and clone selection; Helix  
Research Institute (supported by Japan Key Technology Center etc.)  
and Department of Virology, Institute of Medical Science,  
University of Tokyo.

## FEATURES

## source

1..2585  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2RP4001064"  
/cell\_line="NT2"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2RP4"  
/note="Cloning vector: pME18FL3-mRNA from NT2 neuronal  
precursor cells after 2-weeks retinoic acid (RA)  
induction."  
53..2263  
/note="unnamed protein product"

## CDS

/codon\_start=1  
/protein\_id="BAB55305.1"  
/db\_xref="GI:14042571"  
/translation="MAVRALKLTLLAVAAASQAESEBAGWMTPLDLLFABGTA  
AVAGDPCVLSMERALRSLRALRLRCCTQCAADFPWELDDWSPSPAQAGAA  
ALRDLSPFGGLLRRAACLRCLGPPAAHSLSEMELEPKRSPNYLQVAFKINKLE



KAVAAATFFVGNPEHMEQNLDDYQTMSCVKEADPKDLETPHMQEPRLGVRLYS  
 EQPDAVPHLEAAUQEYFVAYEBCALCEGPDYDGNLYLEYNADLFQALTDHYIQL  
 NCKQCVTELASHSPREKFFDLPUSHYNYLQFAYNIGNYTVAVECAKTYLLFFPND  
 EVMQNLYAYAAAGLEERTSIPGRSAKRYRSLLEKELLFPAYDVGITPFPVDPDS  
 WTPBEVPIKQEQKSEKTRAVRISQIGNLMEKIELTVBEKTKESLDVSRLTREGG  
 PLYBEGISLTWNSKLNSORVMDGVTSDHECOBOLRLTNVAATSGDGYRGOTSPET  
 PNEKPYGVTFKALKLQEGKVLQSAHLVYNTVEKVRIMESVRLDTPLYYSYSL  
 VCRFAIERSQERKDDSPHVDNCLINAEITLVCKEPPAYTRDYSAIILYLAGDFPG  
 CNFPTEDAKITVTRAVQPCQRAVGFSSFENPHGVKAVTRGQRCAIALWFLDPDR  
 SERDRVQADDLVKMLFSPSEMDLSOEPLDAQQGPFPBFAQESUSGSSSKPKDEL"

## ORIGIN

## Alignment Scores:

Pred. No.: 1,896-293 Length: 2585  
 Score: 3846.00 Matches: 733  
 Percent Similarity: 99.59% Conservative: 0  
 Best Local Similarity: 99.59% Mismatches: 3  
 Query Match: 99.38% Indels: 0  
 DB: 9 Gaps: 0

US-10-045-815-4 (1-736) x AK027697 (1-2585)

QY	1	MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer	20
DB	53	ATGGCGGTACCGCGGTGAAGCTGTGACCACTGCTGGCTGTGGCGGCTGCCTCC	112
QY	21	GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla	40
DB	113	CAGCCGAGGTGAGTCCGAGCGCAGGATGGGCGATGTGACCGCTGATCTGCTTTCGCC	172
QY	41	GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg	60
DB	173	GAGGGGACCGAGCTACCGCGCGGGGACTGGCGCGGGTGGTCTCTGAGCATGGAAACGG	232
QY	61	AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla	80
DB	233	CGCTGCGCTCCCGGCGAGCCCTCGCGCCCTTCGCTGCGCTCGCGACCCAGTGTGCC	292
QY	81	AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly	100
DB	293	GCCGACTTCCCGTGGAGCTGACCCCGACTGGTCCCCCGCCGCGCCAGGCTCGGGC	352
QY	101	AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu	120
DB	353	CGCCGCGCCCTCGCGGACCTGAGCTTCTCGGGGCGCTTCGCGTCCGCTGCTGCTG	412
QY	121	ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe	140
DB	413	CGCGCTGCTCGGGCGCGCGCGCCCTCGCTCAGCAGAGAGATGGAGCTGGAGTTC	472
QY	141	ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu	160
DB	473	CGCAAGCGGAGCCCTCAACTACTCTCAGCTCGCCTACTTCAAGATCAACAAGTTGGAG	532
QY	161	LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln	180
DB	533	AAAGCTGTGTGAGCAGACACACCTTCTTGTGGGCAATCTTGAGCACATGGAAATGCAG	592
QY	181	GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu	200
DB	593	CAGAACCTTAGACTATTACCAACCACTGTCTGAGTGAAGGAGGCGCACTTCAAGGATCTT	652
QY	201	GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln	220
DB	653	GAGACTCAACCCCATATGCAAGAAATTTGAGCTGGAGTGGAGTCTACTCAGAGGAACAG	712
QY	221	ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu	240
DB	713	CCACAGGAAGCTGTGCCCCACTAGAGGCGGCGCTGCAAGATACCTTTGTGGCCTATGAG	772
QY	241	GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr	260
DB	773	GAGTGGCTGCTCTCGGAGGGGCCCTATGACTACGATCGCTACACCTTGTAGTAC	832

QY	261	AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln	280
DB	833	AACGCTGACCTTTCAGGCCCATCACAGATCATTAATCATCCAGTCTCACTCACTGAAGCAG	892
QY	281	AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu	300
DB	893	AACTGTGTCCAGGAGCTTCTCCACCCCAAGTCAGAGAGCCCTTTGAAGACTTCTCT	952
QY	301	ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla	320
DB	953	CCATCGCATATAATATCTGCAGTTTGGCTACTATAACATTGGGAATTATACACAGGCT	1012
QY	321	GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn	340
DB	1013	GTTGAATGTGCCAAGACCTATCTCTCTCTTCCCAATGACGAGGTGATGAACCAAAAT	1072
QY	341	LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu	360
DB	1073	TTGGCTATTATGAGCTATGCTTGGAGAAGAACACACAGATCCATCGGCCCGGTGAG	1132
QY	361	SerAlaLysGluTyrArgGlnArgSerLeuLeuLysGluLeuLeuPhePheAlaTyr	380
DB	1133	AGTGCCAAGAGTACCGACAGCAAGCTTACTTGAAAAAGAACTCTTTCTTCGCTTAT	1192
QY	381	AspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIlePro	400
DB	1193	GATGTTTTTGGAAATTCCTTTGTGGATCCGGATTCATGGACTCCAGAGAGAGTGATTC	1252
QY	401	LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu	420
DB	1253	AAAGATTTCAGAGAGAAACAGAAAGTCAGAACGGGAAACAGCGTACGATCTCCAGGAG	1312
QY	421	IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu	440
DB	1313	ATTGGAACTTATGAGGAAATCGAGACCTTGTGGAGAGAGAACCAAGAGTCACTG	1372
QY	441	AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr	460
DB	1373	GATGTGAGCAGACTGACCCGGGAAGTGGCCCTGCTGTATGAAGGATCAGTCTCACC	1432
QY	461	MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp	480
DB	1433	ATGAATCAAACTCCTGAATGGTTCCAGCGGGTGGTGGAGCGGCTAATCTCTGAC	1492
QY	481	HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr	500
DB	1493	CACGAGTGTGAGGAGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGCTAC	1552
QY	501	ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys	520
DB	1553	CGGGCTCAGACTCCCCACATCTCCCAATGAAAGTTCTATGGTGTCACTGTCTTCAA	1612
QY	521	AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn	540
DB	1613	GCCTCAAGCTGGGCAAGAGGCAAGTTCTCTGACAGAGTGGCCCTGTACTACAC	1672
QY	541	ValThrGluLysValArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr	560
DB	1673	GTGCGGAGAAAGGTGCGCGCATCATGAGTCTTCCGCTCGATACGCCCTCTCTAC	1732
QY	561	PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys	580
DB	1733	TTTTCTCTCTCATCTGTGTGTGCGCACTGCCATCGAAGAGGTCCAGGAGAGGAG	1792
QY	581	AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys	600
DB	1793	GATGATAGTCACTCCAGTCCAGTCCAGCAACATGTCATCTCTGAATGCCAGACCCCT	1852
QY	601	ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly	620
DB	1853	GTCAAAGAGGCCCCAGGCTACACCTTCCGCGACTACAGCGCCATCTTACATAATGGG	1912





DB 1036 TTGGCCCTATTATGACAGCTATGCTTGGAGAGAACACACAGATCCATCGGCCCGCTGAG 1095  
QY 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhePheAlaTyr 380  
DB 1096 AGTCCCAAGGAGTACCGACAGCGAAGCTTACTGGAAAGAACTGCTTTCTTCGCTTAT 1155  
QY 381 AspValPheGlyTyrProPheValAspProAspSerTyrThrProGluGluValPro 400  
DB 1156 GATGTTTGGAAATCCCTTTGGATCCGATTCATGGACTCCAGAGAGATGATCC 1215  
QY 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgLysSerGlnGlu 420  
DB 1216 AAGAGATTCCAGAGAGAAACAGAGATCAGAACCGGAAACAGCCGTACGATCTCCACGAG 1275  
QY 421 IleGlyAsnLeuMetLysGluLeuThrLeuValGluGluLysThrLysGluSerLeu 440  
DB 1276 ATTGGAACTTATGAAGAAATCGAGACCTTGTGGAAGAGAACACCAAGAGTCACTG 1335  
QY 441 AspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyLysLeuThr 460  
DB 1336 GATGTGACGAGACTGACCCGGAAGGTGCCCTCTGTATGAGGATCAGTCTCACC 1395  
QY 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValMetAspGlyValLysSerAsp 480  
DB 1396 ATGAACCTCCAAACTCCTGAATGTTCCACAGCGGTGTGTGATGGACGGGTAATCTCTGAC 1455  
QY 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
DB 1456 CACGAGTGTGAGAGCTGACAGACTGACCAATGTCGAGCAACCTCAGGAGATGGCTAC 1515  
QY 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
DB 1516 CGGGTGCAGACCTCCCACTATCCCAATGAAGTTCTATGGTGTCTACGTCTTCAA 1575  
QY 521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
DB 1576 GCCCTCAAGCTGGGCAAGAGGCAAAAGTTCTCTGCAGAGTGCCCACTGTACTACAAC 1635  
QY 541 ValThrGluLysValArgArgLeuMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
DB 1636 GTGACGGAGAGGTGCGCGCATATGAGTCTTCTTCCCTCGGATACCCCTCTAC 1695  
QY 561 PheSerTyrSerHisLeuValCysArgThrAlaLeuGluValGlnAlaGluArgLys 580  
DB 1696 TTTTCTACTCTCTCTGTGTGCGCACTGCCATCGAAGAGTCCAGGCGAGAGGAAG 1755  
QY 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
DB 1756 GATGATAGTATTCAGTCCAGTCCAGTGGACACTGTCATCTGTAATCCGAGACCTCGTGTGT 1815  
QY 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
DB 1816 GTCAAGAGCCCCAGCCTACCTTCGCGACTACAGCCCATCTTTTACCTAAATGG 1875  
QY 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
DB 1876 GACTTCGATGGCGAAACTTTATTTTCACTGAATCGGATCCAGACCTGAAACCCACATGGA 1935  
QY 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
DB 1936 GTGCAGCCTCAGTGTGAAGAGCGGTGGATCTCTTCAGGCATCGAAGACCCACATGGA 1995  
QY 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuAspPro 680  
DB 1996 GTGAAGGCTGTACACAGGGGCGAGCGGTGTGCCATCGCTCTGTGGTTTCACTCGACCT 2055  
QY 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
DB 2056 CGACACAGCGAGCGGACAGGTCAGGAGATGACCTGTGTGAAGATGCTCTTCAGCCCA 2115  
QY 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProGluPro 720  
DB 2116 GAAGAGATGACCTCTCCAGGAGCAGCCCATGATGCTCCAGAGGCGGCCCGCAACCT 2175

QY 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
DB 2176 GCACAGAGTCTCTCTCAGGCACTGATCGAAGCCCAAGATGAGCTA 2223

RESULT 10  
BD156505 2524 bp DNA linear PAT 17-JAN-2003  
LOCUS Primer for synthesizing full-length cDNA and use thereof.  
DEFINITION  
ACCESSION BD156505  
VERSION BD156505.1 GI-27862263  
KEYWORDS JP 2002191363-A/11348.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2524)  
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,  
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.  
TITLE Primer for synthesizing full-length cDNA and use thereof  
JOURNAL Patent: JP 2002191363-A 11348 09-JUL-2002;  
HELIX RESEARCH INSTITUTE  
COMMENT OS Homo sapiens (human)  
PN JP 2002191363-A/11348  
PD 09-JUL-2002 JP 2002280990  
PF 28-JUL-2000 JP 2002280990  
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU  
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
PI KEIICHI NAGAI, TETSUJI OTSUKI  
PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC  
10, C12P21/02, C12Q1/69//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC  
Primer for synthesizing full-length cDNA and use thereof FH Key  
Location/Qualifiers  
(16) .. (2223).  
FT CDS Location/Qualifiers

FEATURES  
source

1..2524  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

ORIGIN

Alignment Scores:

Pred. No.: 2,63e-293 Length: 2524  
Score: 3844.00 Matches: 732  
Percent Similarity: 99.59% Conservative: 1  
Best Local Similarity: 99.46% Mismatches: 3  
Query Match: 99.33% Indels: 0  
DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x BD156505 (1-2524)

QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
DB 16 ATGGCGGTACGCGCGCTTGAAGCTGTGACACACTGTGGCTGTGGCGCGCTGCCCTCC 75  
QY 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuPheAla 40  
DB 76 CAAAGCCGAGTGTGAGTCCGAGCAGGATGGGCGCATGTGACGCTGATCTCTTTCGCC 135  
QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60  
DB 136 GAGGAGACCGCAGCCTACGCGCGCGGACTGTGGCGGGTGTCTGTGAGCATGGAACGG 195  
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
DB 196 GCGCTGCGCTCCCGGAGAGCCCTCCGCGCCCTTCGCTGGCGCTCCGACCATGATGTGCC 255  
QY 81 AlaAspPheProTyrGluLeuAspProAspTyrSerProAlaGlnAlaSerGly 100  
DB 256 GCCGACTTCCGTTGGAGTCTGACCCCGACTGTGTCCTCCCGAGCCCGCCGCTCGGGC 315

101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
102 |||||  
316 GCCGCCCTCGCGACCTGAGCTTCTTCGGGGGCTTCTCGCTGCGCTGCTGCTGCTG 375  
121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
122 |||||  
376 CGCGCTGCTCGGGCGCGCGCGCGCCCTCGCTCAGCGAAGAGATGAGCTGGAGTTC 435  
141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysLysLeuLeuGlu 160  
142 |||||  
436 CGCAAGCGGAGCCCTCAACTACTCTGAGTGGCTGCTTCAAGATCAACAAGTGGAG 495  
161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180  
162 |||||  
496 AAGCTGTGCTGCGAGCACACACCTTCTTGGTGGCAATCTCTGAGCACATGGAAATGCAG 555  
181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
182 |||||  
556 CAGAACCTAGACTATTACCAACCATGTCTGGAGTGAAGGCGGCTTCAAGGATCTT 615  
201 GluThrGlnProHisMetGlnGluPheArgGluGlyValArgLeuTyrSerGluGluGln 220  
202 |||||  
616 GAGACTCAACCCCATATGCAAGATTTGCACTGGAGTGGAGCTCTACTCAGAGGAACAG 675  
221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
222 |||||  
676 CCACAGGAGCTGTGCCCACTAGAGCGCGCTGCAAGAATACTTTGTGGCTATGAG 735  
241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
242 |||||  
736 GAGTGGCTGCTGCGAAGGCGCTTATGACTACGATGGCTCAACTACCTTGGATGAC 795  
261 AsnAlaAspLeuPheGlnAlaThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
262 |||||  
796 AACCTGACCTTCTCCAGGCCATCACAGATCATTACATCCAGGTCTCTCAACTGAAGCAG 855  
281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
282 |||||  
856 AACTGTGTCCGAGCTTGTCTCCACCAAGTCGAGAGAGCCCTTTGAAGACTTCTCTC 915  
301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGluAla 320  
302 |||||  
916 CATCGCATTAATAATATCTGAGTTTGTCTACTATATACATGGGAATATACACAGCT 975  
321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
322 |||||  
976 GTTGAATGTGCCAAGACCTATCTTCTCTTCCCAATGACGAGGTGATGAACCAAAAT 1035  
341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
362 |||||  
1036 TTGGCTATTATGCGACTATGCTTGGAGAAGAACAACACAGATCCATCGGCCCGTGAG 1095  
361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhePheAlaTyr 380  
382 |||||  
1096 AGTCCAGAGATACCGACAGGAGCTTCTGGAAGAAGAACTCTTTTCTGCTTAT 1155  
381 AspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIlePro 400  
402 |||||  
1156 GATGTTTTTGAATTCCTTTTGGATCGGATTCATGACTCCAGAAAGATGATTCCTCC 1215  
401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgLysSerGlnGlu 420  
422 |||||  
1216 AAGAGATTGCAGAGAAACAGAAAGTCAGAACCGGGAACAGCGCTACGCACTCTCCAGGAG 1275  
421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
442 |||||  
1276 ATTGGGAACCTTATGAAGAAATCGAGACCTTGTGGAAGAGAAAGAACCAAGGAGTCACTG 1335  
441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluLysLeuThr 460  
462 |||||  
1336 GATGTGAGCAGACTGACCCGGGAAGGTGGCCCCCTCTGTATGAGGAGCATCACTCTACC 1395

461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
482 |||||  
1396 ATGAACCTCAAACTCCTGAATGGTTCCAGAGGGGTGGTATGGAGCGGCTAATCTCTAC 1455  
481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
482 |||||  
1456 CACGAGTGTACGAGACTGCAGAGACTGACCAATGTGGCAGACACCTCAGGAGATGGTAC 1515  
501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
522 |||||  
1516 CGGGGTGACAGCTCCCACTACTCCCAATGAAAGTTCTATGGTGTCTGCTCTCTCAA 1575  
521 AlaLeuLeuLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
542 |||||  
1576 GGCCTCAAGCTGGGCAAGAGCAAGTTCTCTGCAGAGTGCCACCTGTACTACTACAC 1635  
541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
562 |||||  
1636 GTGACGGAGAGGTGGGGCATCATGGAGTCTTACTTCCGCTGGATACGCCCTCTAC 1695  
561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
582 |||||  
1696 TTTTCTACTCTCATCTGGTGTGGCGCACTGCATCGAAGAGGTCCAGGACAGAGGAAG 1755  
581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
582 |||||  
1756 GATGATGATCATCCAGTCCAGTGGACACTGCATCTCTGAATGCCGAGACCTCGTGTCT 1815  
601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
622 |||||  
1816 GTCAAGAGAGCCCGACCTACACCTTCCGCACTACAGCGCATCTTTTACCTAAATGG 1875  
621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
642 |||||  
1876 GACTTCGATGGCGGAACCTTTTATTTCACTGAACTGATGATGCCNAGACCTGTGAGCGAG 1935  
641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
662 |||||  
1936 GTGACGCTCAGTGTGGAAGCGCTGGGATTTCTTCCAGCACTGAAACCCACATGGA 1995  
661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680  
682 |||||  
1996 GTGAAGCTGTCAACAGGGGCGCGCTGTGCCATCGCTGCTGGTTTACCTTGGACCT 2055  
681 ArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMetLeuPheSerPro 700  
682 |||||  
2056 CGACACAGCGAGCGGACAGGCTGCAGGAGATGACCTGTGGAAGATGCTTCTTCAGCCCA 2115  
701 GluGluMetAspLeuSerGlnGluProLeuAspAlaGlnGlnGlyProProGluPro 720  
722 |||||  
2116 GAAGAGATGGACCTTCTCCAGGAGAGCCCATGGATGCCAGAGGCGGCCCGCCGAACT 2175  
721 AlaGlnLeuSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
737 |||||  
2176 GCACAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2223

RESULT 11  
AK027680 2524 bp mRNA linear PRI 01-AUG-2002  
LOCUS Homo sapiens cDNA FLJ14774 fis, clone NT2RP4000051, weakly similar  
DEFINITION to SYNAPTONEMAL COMPLEX PROTEIN SC65.  
ACCESSION AK027680  
VERSION AK027680.1 GI:14042539  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
REFERENCE  
AUTHORS  
Isegai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,  
Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y.,  
Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,  
Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Sasaki,N.



```
Db 1336 GATGTGACGACATGACCCGGAGGTGGCCCTGCTGATGAAGCATCATCTCCACC 1395
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480
Db 1396 ATGAACCTCCAACTCTCTGAATGTTCCAGCGGTGTGATGACGGCGTAATCTCTGAC 1455
Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaIleThrSerGlyAspGlyTyr 500
Db 1456 CACGAGTGTGAGGAGTGCAGAGACTGACCAATGTGCAGCAACCTCAGAGATGGCTAC 1515
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520
Db 1516 CGGGGTGAGACTCCCCACATCTCCCAATGAAAGTTCTATGTTGTCACCTGCTTCAA 1575
Qy 521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540
Db 1576 GCCCTCAAGCTGGGGCAAGGCAAGTTCCTCTGCAGAGTGCACCTGTACTACAAC 1635
Qy 541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560
Db 1636 GTGACGAGAAAGTGCAGCGCATCATGAGTCTTACTTCCGCTGGATACGCCCTCTAC 1695
Qy 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580
Db 1696 TTTTCCTACTCTCATCTGTTGCGGCACTGCCATCGAAGAGTCCAGGCAGAGGAAG 1755
Qy 581 AspAspSerHisProValHisValAspAsnCysIleLeuLeuAlaGluThrLeuValCys 600
Db 1756 GATGATGATCATCCAGTCCAGTGCACAACTGCATCTGATGCGGACCTCGTGTGT 1815
Qy 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620
Db 1816 GTCAAGAGAGCCCCAGCCCTACACCTTCGCGACATACAGCGCCATCTTACCTAATGGG 1875
Qy 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640
Db 1876 GACTTCGATGCGGAAACTTTTATTTCACTCAACTGATGATGCAAGCCGTGACGCGCAGAG 1935
Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660
Db 1936 GTGACGCTCATGTGTGGAAGACCGGTGGATTCCTTCAGGACATGAAACCCACATGGA 1995
Qy 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680
Db 1996 GTGAAGGTGTGTACACAGGGGCGAGCGCTGTGCCATCGCCCTGTGTTTCACTGAGCCT 2055
Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700
Db 2056 CGACACAGCAGCGGGACAGGCTGCAGGCAGATGACCTGGTCAAGATCTCTTCAGCCCA 2115
Qy 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProGluPro 720
Db 2116 GAAGAGATGGACCTCTCCAGAGAGCAGCCCATGGATGCCAGCAGGGGCCCCCGGACCT 2175
Qy 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736
Db 2176 GCACAAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2223

RESULT 12
AF097431
LOCUS      Homo sapiens GROSI-S protein mRNA, complete cds.
DEFINITION
ACCESSION AF097431
VERSION    AF097431.1 GI:11127635
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 2829)
AUTHORS    Kaul, S.C., Sugihara, T., Yoshida, A., Nomura, H. and Wadhwa, R.
TITLE      GroS1, a potential growth suppressor on chromosome 1: its identity
```

```
to basement membrane-associated proteoglycan, leprecan
Oncogene 19 (32), 3576-3583 (2000)
20406537
REFERENCE  2 (bases 1 to 2829)
AUTHORS    Wadhwa, R., Sugihara, T. and Kaul, S.
TITLE      Direct Submission
JOURNAL    Submitted (06-OCT-1998) Group 4, Chugai Research Institute of
Molecular medicine, 153-2 Nagai, Nihari, Ibaraki 300-41, Japan
FEATURES
source
1..2829
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="1"
/tissue_type="testis"
52..1143
/note="member of protein B family"
/codon_start=1
/product="GROSI-S protein"
/protein_id="AAG31018.1"
/db_xref="GI:11127636"
/translation="MAVRALKLLTLLAVVAASQAQVESEAGMGVTPDLLFASGTA
AYARDGPGVLSMERALRSRAALRLRCKTCAADFPELDDPMSFPAQASGAG
ALRDLSPFGGLLRRAACLRRCIGPPAAHSLSEMELEFRKSPYINLYQVAFKINKLE
KAVAAAHTPFVGNPEHMQNDLYOTMSGVKEADFKDLETPQHMQRFLGRLYSE
EOQAVPHEALAOEYFYAYEBECALCEGPDYDYNLEYNADLFOAITDHYIQVL
NKKQCVTELASHPSREKFEFDLPShYVNLQFAYVNIQNTQACCAKTYLLFFPND
EYVQNQLAYAAVLGEHTRSGPREQGT"
ORIGIN
Alignment Scores: 3.43e-288 Length: 2829
Pred. No.: 3780.00 Matches: 733
Score: 90.41% Conservative: 2
Percent Similarity: 90.16% Mismatches: 1
Best Local Similarity: 97.67% Indels: 77
Query Match: 9 Gaps: 2
DB:
US-10-045-815-4 (1-736) x AF097431 (1-2829)
Qy 1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValValAlaAlaSer 20
Db 52 ATGGCGGTACGCGGTGAGCTGTGACCACTGCTGGCTGCTGCGCGCTGCTCC 111
Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuPheAla 40
Db 112 CAAGCCGAGTCCAGTCCGAGGCGAGGATGGGCGATGTCAGCGCTGATCTCTTCGCC 171
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60
Db 172 GAGGGACCCGACGCTACGCGCGCGGAGCTGGCCCGGGGTGGTCTCTGAGCATGGACGG 231
Qy 61 AlaLeuArgSerArgAlaAlaLeuAlaLeuArgLeuArgCysArgThrGlnCysAla 80
Db 232 CGCGTGGCTCCCGGGCAGCCCTCCGCGCCCTTCGCTGGCTGCGCCAGCCAGTGTGCC 291
Qy 81 AlaAspPheProTyrGluLeuAspProAspTyrSerProSerProAlaGlnAlaSerGly 100
Db 292 GCGGACTTCCCGTGGGAGCTGGACCCCGCATGCTGCCCGCCGCGCCAGGCTCGGGC 351
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyLeuLeuArgArgAlaAlaCysLeu 120
Db 352 GCGGCGCCCTGCGCGACCTTCTCGGGGGCTTCTGGCTGCGCGCTGCTGCTGCTG 411
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140
Db 412 CGCGCTGCTCGCGCGCGCGCGCCCTCGCTCAGCGAAGATGGAGCTGGAGTTC 471
Qy 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160
Db 472 CGCAAGCGGAGCCCTCAACTACCTCAGTCCGCTTCTCAAGATCAACAAGTTGGAG 531
```

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source

QY 161 LysAlaValAlaAlaHisThrPhePheValGlyVasnProGluHisMetGluMetGln 180  
 Db 532 AAAGCTGTTGTCGACGACACACACCTCTCTCTGCGCAATCCTGAGCACATGGAATGCG 591  
 QY 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
 Db 592 CAGAACCTAGACTATTACCAACCAATGCTCTGAGTGAAGAGGCGGCACTTCAAGATCTT 651  
 QY 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220  
 Db 652 GAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTGGCACTTACTCAGAGAAACAG 711  
 QY 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
 Db 712 CCACAGGAAGCTGTGCCCCACCTAGAGGCGGCGCTGCAGAAATACTTTGTGGCCTATGAG 771  
 QY 241 GluCysArgAlaLeuGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
 Db 772 GAGTGGCGTCCCTCTGCGAAGGCGCTATGACTAGATGGCTACAACTACCTTGGATG 831  
 QY 261 AsnAlaAspPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
 Db 832 AACGCTGACCTCTCCAGGCCATCACAGATATTATCATCCAGGTCCTCAACTGTAAAGCAG 891  
 QY 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
 Db 892 AACTGTGTCCAGGAGCTGTCTCCACCAAGTCCAGAGAGGCGCTTTGAAGACTTCCTC 951  
 QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyVasnTyrThrGlnAla 320  
 Db 952 CCATCGCATTAATATCTGAGTTGCTTACTATATACATTTGGAAATTTACACAGCT 1011  
 QY 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
 Db 1012 GGTGAATGTGCCAAGACCTATCTCTCTCTCCCAATGACGAGGTGATGAACAAAT 1071  
 QY 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluHisThrArgSerIleGlyProArgGlu 360  
 Db 1072 TTGGCTATTATGAGCTATGTTGGAGAGACACACACATCATCGGCCCGCTGAG 1131  
 QY 361 -----SerAl 362  
 Db 1132 CAGGGCACCTAGGAAAGATGTGACCCCGAAAGTACTCAGTTCCCTCGCCCTGAGTGC 1191  
 QY 362 alysgluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaTyrAspVa 382  
 Db 1192 CAAGAGTACCCAGCAGGAGCTACTGGAAGAAAGACTGCTTTCTTCGCTTATGATGT 1251  
 QY 382 lPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIleProLysAr 402  
 Db 1252 TTTTGGAAATCCCTTTGTGGATCGGATTCATGAGTCCAGAGAAATGATTCCCAAGAA 1311  
 QY 402 gheuGlnGluLysGlnLys----- 408  
 Db 1312 ATTGCAGAGAAACAGAAAGTGGAGCCTTTGAAGAAACTGCATGTTGGATCAGTCTGATG 1371  
 QY 408 ----- 408  
 Db 1372 AAGCACTTGAGCTTCTTTGACCCAGGCGAGATGTGAACCTCTGCGCAAGGGTGGCGAGGT 1431  
 QY 408 ----- 408  
 Db 1432 CCAGTTTGGGAAGTGGGGTGGAGCCCGAGGGCTGGCTGGAATGAGTCACTCAGAGCGG 1491  
 QY 409 -----SerGluArgGluThrAlaValArgIleSerGlnGluIleGlyAsnLe 424  
 Db 1492 TTGTGCTCATAGTTCAGAACGGGAAACAGCGGTACGATCTCCAGAGATTTGGAACTT 1551  
 QY 424 uMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeuAspValSerAr 444  
 Db 1552 TATGAAGGAATCGAGACCTTTGTGAAGAGAGAACCAAGAGTCACTGGATGTGAGCAG 1611  
 QY 444 gLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThrMetAsnSerly 464

Db 1612 ACTGACCCGGGAAGGTGGCCCTCTGTGTATGAAGCATCAGTCTCACCATGAACTCCAA 1671  
 QY 464 sLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAspHisGluCysGl 484  
 Db 1672 ACTCTGTAATGGTTACCAGCGGTGGTGTATGGACGCGTAATCTCTGACACAGTGTCA 1731  
 QY 484 nGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyVasnGlyTyrArgGlyGlnTh 504  
 Db 1732 GGAGCTGCAGAGACTGACCAATGTGGCAGCACTCAGAGATGGCTACGGGGTGCAGAC 1791  
 QY 504 rSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLysAlaLeuLysLe 524  
 Db 1792 CTCCCACTACTCCCAATGAAAAGTTCTATGGTGTCACTGTCTTCAAAGCCCTCAAGCT 1851  
 QY 524 uGlyGlnGluGlyValProLeuGlnSerAlaHisLeuTyrTyrAsnValThrGluLys 544  
 Db 1852 GGGGCAAGAGGCAAGATTCCTCTGAGAGTGGCCACCTGTACTACAGGTGACGAGAA 1911  
 QY 544 sValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyrPheSerTyrSe 564  
 Db 1912 AGTGGCGGCATCATGGAGTCTCTTCCGCTGGATACGCCCTCTACTTTTCTACTC 1971  
 QY 564 rHisLeuValCysArgThrAlaIleGluValGlnAlaGluArgLysAspAspSerHi 584  
 Db 1972 TCATCTGGTGTGGCGCACTGCCATCGAAGAGGTCCAGGAGAGAGGATGATAGTCA 2031  
 QY 584 sProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCysValLysGluPr 604  
 Db 2032 TCAGTCCAGTGGCAACTGATCTCTGATGCGGAGACCTCTGCTGTGTCTCAAGAGCC 2091  
 QY 604 oProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGl 624  
 Db 2092 CCCAGCCTACACCTTCGCGCACTACAGGCCATCTCTTACCTAAATGGGCACTTCGATGG 2151  
 QY 624 yGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGluValGlnProGl 644  
 Db 2152 CGGAACTTTTATTTCTACTGAATGGATGCCAAGACCGTGACGGAGAGGTGCAGCTCA 2211  
 QY 644 nCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGlyValLysAlaVa 664  
 Db 2212 GTGTGAAGAGCGTGGGATCTCTTCAGGCCACTGAAAAACCCACATGGAGTGAAGCTGT 2271  
 QY 664 lThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspProArgHisSerGl 684  
 Db 2272 CACAGGGGCGAGCGGTGTGCCATCGCTGTGGTTCACCTGGACCTCGACACAGCGA 2331  
 QY 684 uArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerProGluGluMetAs 704  
 Db 2332 CGCGGACAGGGTGCAGGAGATGACCTGTGTGAAGATGCTCTTCAGCCCAAGAGAGATGGA 2391  
 QY 704 pleuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluProAlaGlnGluSe 724  
 Db 2392 CCTCTCCAGGAGAGCGCCCTGGATGCCAGCAGGGCCCCCCCCCAACCTGCACAAAGTGC 2451  
 QY 724 rLeuSerGlySerGluSerLysProLysAspGluLeu 736  
 Db 2452 TCTCTCAGGCACTGAATCGAAGCCCAAGGATGAGCTA 2488

## RESULT 13

AX880254  
 LOCUS AX880254  
 DEFINITION Sequence 15159 from Patent EP1074617.  
 ACCESSION AX880254  
 VERSION AX880254.1 GI:40034990  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1  
 Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J., Ishii, S., Sugiyama, T., Wakamatsu, A., Negai, K. and Otsuki, T.

TITLE Primers for synthesizing full-length cDNA and their use  
JOURNAL Patent: EP 1074617-A 15159 07-FEB-2001;  
Research Association for Biotechnology (JP)

FEATURES  
source  
1. .2993  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
42. .2456  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAE90869.1"  
/db\_xref="GI:40034991"  
/translation="MAVRALKLTLTLLAVAAASQAESEAGWMTVDLLFABGTA  
AVARGDWPGLVSMERALSRAALRLRCRTQCAADFPWELDPWSPSPQAAGAA  
ALRDLSPFGGLLRCLPAAHSLSEMELEFRKSPYNYLQVAYFKINKLE  
KAVAAHPTVPGNPEHMEWQNDLYQVMSGVKEADFKLETPHMQEPRIGVRLYSE  
EQQEAVPHLEALQYFVAYEBCRALCEPDYDGYNYLEVADLFOALTDHYLOVL  
NKQNCVTELAHSPSEKPEFPLSHNYLQFYNYNIGNYQAVCAKTYLLFPND  
EVNNQLYAALMGHEHRSIGTPRESAKEYRQSLLEKELUFFAYDVFQIPFVDPDS  
WTPGEVTPKQKQKSERETAVRIISOEIGNLMEKTELVEEKTESLDSVRLREGG  
PLLYEGISLTNKLQSGVQVMDGVSIDHEQELQRLTNVAATSGDGYRGQTSPT  
PNKFGVTVFKALQGOEKVPLQSAHLVYNTVEKVRIMESYFRLDTPLYPSVSHL  
VCSTAEEVQAEKDSHPVHVNDILNAETLVCKVEPPAYTRDYSAILYLNGDDPG  
GNFYFELDKTVAEVQPOCGRAVGFSSGTENPHGVKAVTRQRCALALWTLDPRH  
SERVRAARGESNWCDDPPERPWFALFPKSHCQWLHERSTWDTSSNALSLMSHC  
LVLPFGPAVNGIQVKEVKGTSDAEFLVPSLGPSTSAVLFORVGPAGKEMSLGPNLUPC  
PLGSSS"

ORIGIN

Alignment Scores:  
Pred. No.: 6,14e-284 Length: 2993  
Score: 3726.50 Matches: 730  
Percent Similarity: 83.33% Conservative: 0  
Best Local Similarity: 83.33% Mismatches: 6  
Query Match: 96.29% Indels: 141  
DB: 6 Gaps: 1

US-10-045-815-4 (1-736) x AX880254 (1-2993)

Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValValAlaAlaSer 20  
Db 42 ATGCGGTACGCGGTTGAAGCTCTGACACACACTGCTGGCTGCTGCGCGCTGCC 101  
Qy 21 GluAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40  
Db 102 CAAGCCGAGGTCGAGTCGAGCGAGGATGGGCGATGGCGCTGATCTGCTCTTCC 161  
Qy 41 GluGlyThrAlaAlaValArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
Db 162 GAGGGGACCGGAGCTACGCGCGCGGGGACTGGCGCGGGGTGGTCTCTGAGCATGGA 221  
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
Db 222 GCGCTGCGCTCCCGGCGAGCCCTCCGCGCCCTTCGCGCTGCGCTCCGCGACCCAGT 281  
Qy 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
Db 282 GCCGACTTCCCGTGGGAGCTGGACCCGACCTGGTCCCGCCGCGCGCGCGCTCGGGC 341  
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
Db 342 GCGCGCGCCCTGCGGACCTGAGCTTCTTCGGGGGCGCTTCGCTGCGCTGCGCTGCTG 401  
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPhe 140  
Db 402 GCGCGCTGCGCTGGGCGCGCGCGCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 461  
Qy 141 ArgLysArgSerProTrpAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
Db 462 CGCAGCGGAGCCCTCTACACTACTCTGAGTCTGCTGCTGCTGCTGCTGCTGCTG 521  
Qy 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180

Db 522 AAAGCTGTGTGTCAGCACACACCTTCTTCTGGGCAATCTGAGCATGGAATGAG 581  
Qy 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyVallysGluAlaAspPheLysAspLeu 200  
Db 582 CAGAACCTAGACTATTACCAACCATGTCTGAGTGAAGAGGCGGACTTCAAGGATCTT 641  
Qy 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGln 220  
Db 642 GAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTGCAGTCTACTCAGAGAACAG 701  
Qy 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
Db 702 CCACAGGAGCTGTGCCCCACTAGAGCGGCGCTGCAAGAAATCTTTGTGGCTATGAG 761  
Qy 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
Db 762 GAGTCCGCTGCGCTCTCGAAGGCGCTTATGACTACGATGCTACCACTACCTTGAGTAC 821  
Qy 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
Db 822 AACGCTGACCTCTCCAGGCCATCAGATCATATACATCCAGGTCTCACTGTAGAGAG 881  
Qy 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
Db 882 AACTGTGTCACGAGCTTGTCCACCAAGTCGAGAGAGCCCTTTTGAAGACTTCTCTC 941  
Qy 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
Db 942 CCATGCGCATTAATATATCTGCGATTTGCTACTATAACATTTGGGAATTTATACAGGCT 1001  
Qy 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
Db 1002 GTTGAAGTGGCCAGACCTATCTTCTTCTTCCCAATGACGAGGTGATGAACCAAAAT 1061  
Qy 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
Db 1062 TTGGCTATTATGCGCTATGCTTGGAGAGAAACACACAGATCCATCGCCCGCTGAG 1121  
Qy 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr 380  
Db 1122 AGTGCCAGAGGATGACGAGCGAAGCTTCTGGAAGAAAGAACTGCTTTCTTCGCTTAT 1181  
Qy 381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400  
Db 1182 GATGTTTTTGGAAATTCCTTTTGGATCCGATTCATGGACTCCAGGAGAACTGATCC 1241  
Qy 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
Db 1242 AAGAGATTGCAAGAGAAACAGAAAGTCAGAACCGGAAACAGCCGTACGCTATCTCC 1301  
Qy 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluLysThrLysGluSerLeu 440  
Db 1302 ATTGGGAACCTTATGAGGAATTCGAGACCTTTGTGGAAGAGAGAGACCAAGGAGT 1361  
Qy 441 AspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
Db 1362 GATGTGAGCAGACTGACCCGGGAAGGTGGCGCCCTGCTGTATGAAGGCAATCAGTCT 1421  
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
Db 1422 ATGAATCCAAACTCTCTGAATGTTCCCGCGGTGGTGGATGGACGCGCTAATCTCTGAC 1481  
Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
Db 1482 CACGAGTGTGAGGAGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGCTAC 1541  
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
Db 1542 CGGGGTGAGACCTCCCAATGAGTATCTGAGTCTGCTGCTGCTGCTGCTGCTGCTG 1601  
Qy 521 AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540



Db 1602 GCCTCAAGCTGGGCAAGAGCAAGTTCTCTGAGAGTGCCACCTGTACTACAAAC 1661  
QY 541 ValThrGluValArgArgileMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
Db 1662 GTACGAGAGAGTGGCGGCATCATGGAGTCTACTTCCGCTGGATACGCCCTCTAC 1721  
QY 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGluAlaGluArgLys 580  
Db 1722 TTTTCTCTCTCATCTGTGTGGCGCACTGCGCATCGAAGAGGTCCAGCGAGAGGAAG 1781  
QY 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
Db 1782 GATGATAGTATCATCCAGTCCAGTGCACAACTGCATCTCTGATCGGAGACCTCGTGTGT 1841  
QY 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
Db 1842 GTCAAAGAGCCCCCAGCCTACACCTTCCGCGACTACAGCGCATCTTTACCTAAATGGG 1901  
QY 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
Db 1902 GACTTCGATGGCGGAACTTTATTTCACTGACTGATGCCAGACCGTGCAGCGCAGAG 1961  
QY 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
Db 1962 GTGCAGCCTCAGTGTGAAGAGCGCGTGGGATTTCTTTCAGGCACTGAAAAACCCACATGA 2021  
QY 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680  
Db 2022 GTGAAGGCTGTACACAGGGCGCGCGCTGTCATTCGCTGTGTTTCCCTGACCCCT 2081  
QY 681 ArgHisSerGluArg----- 685  
Db 2082 CGACACAGCGAGCG-GGTGAGAGCAGCTCGAGCGGTGAGAGCAGCTGGTCTGTGTGA 2140  
QY 685 ----- 685  
Db 2141 CCGGTTCACAGAGCGCCCTGGTTGGCTTTCTCTTCTCCCAAAATCCCATGTCAGTGCT 2200  
QY 685 ----- 685  
Db 2201 GAGACACGAAGAGGACCTTGGACACCAAGCTCCCAAGCCCTGTCAATATGGTCAACATTG 2260  
QY 685 ----- 685  
Db 2261 CTTTGTCTCCTTGGGCTGCTGTGAACGGGATCCAGTGGGGAAGAGGTCAAGACAGG 2320  
QY 685 ----- 685  
Db 2321 GAGCGATGCTGAGTCTTGGTTTCTCTCTGGGCGCCCACTTCAGCTGTCTCTTTCCAGAG 2380  
QY 685 ----- 685  
Db 2381 AGTAGGACCTGCTGGGAAGGAGATGAGCTGGGGCCATTAAGGAACCTTCTTGTCTCCCT 2440  
QY 685 ----- 685  
Db 2441 GGGAGTAGCAGCTGAGAGATAGCGAGTGTCTGGAGCGGAGGCTCTCTGAATGGGCGAG 2500  
QY 686 ----- AspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
Db 2501 GGTTTGTCTTGTGAGGACAGGCTGAGCTGGGGCCATTAAGGAACCTTCTTGTCTCCCT 2560  
QY 701 GluGluMetAspLeuSerGlnGluProLeuAspAlaGlnGlnGlyProProGluPro 720  
Db 2561 GAAGAGATGTCTCTCCAGAGACAGCCCTGTGATGCCAGAGGCGCCCGGAACT 2620  
QY 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2621 GCACAAGATCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2668

RESULT 14

BD158280

LOCUS

BD158280 2993 bp DNA linear PAT 17-JAN-2003

DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Primer for synthesizing full-length cDNA and use thereof.

BD158280

BD158280.1 GI:27864038

JP 2002191363-A/13123

Homo sapiens (human)

Homo sapiens

Homo sapiens

Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2993)

Osaka, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,

Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.

Primer for synthesizing full-length cDNA and use thereof

Patent: JP 2002191363-A 13123 09-JUL-2002;

HELIX RESEARCH INSTITUTE

OS Homo sapiens (human)

PN JP 2002191363-A/13123

PD 09-JUL-2002

PF 28-JUL-2000 JP 2000280990

PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU

PI SAITO,

PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,

PI KEIICHI NAGAI, TETSUJI OTSUKI

PC

C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/10,

PC, C12P21/02, C12Q1/68, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC

Primer for synthesizing full-length cDNA and use thereof FH Key

Location/Qualifiers

(42)..(2453).

FT CDS

Location/Qualifiers

1..2993

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

ORIGIN

Alignment Scores:

Pred. No.: 6,14e-284 Length: 2993

Score: 3726.50 Matches: 730

Percent Similarity: 83.33% Conservative: 0

Best Local Similarity: 83.33% Mismatches: 6

Query Match: 96.29% Indels: 141

DB: Gaps: 1

US-10-045-815-4 (1-736) x BD158280 (1-2993)

QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaAlaSer 20

Db 42 ATGGCGGTACGCGCGCTGAAGCTGTGACACACTGTGGCTGTGTCGGCGCTGCCCTCC 101

QY 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuPheAla 40

Db 102 CAAGCCGAGGTGAGTCCGAGCGAGATGGGCGATGGTGGCTGATCTCTTCTGCC 161

QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60

Db 162 GAGGCGCGCGCGCTACGCGCGGAGCTGGCGCGGCTGGTGGCTGGATGGAACGG 221

QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80

Db 222 GCGCTGCGCTCCCGGCGAGCCCTCCGCGCCCTTCCGCTGCGCTGCGCGCACCATGTGTC 281

QY 81 AlaAspPheProTyrGluLeuAspProAspTyrPsrProSerProAlaGlnAlaSerGly 100

Db 282 GCGGACTTCCGCTGGGAGCTGGACCCGACTGTGTCGCCCGCGCGCGCGCGCGCC 341

QY 101 AlaGlyAlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgAlaAlaCysLeu 120

Db 342 GCGCGCGCGCTTCCGCGAGCTGAGGCTTCTTCGGGGCGCTTCTGCGCTGCGCTGCGCTG 401

QY 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluMetGluLeuGluPhe 140

Db 402 GCGCGCTCTCGGCG 461



141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
142 CGCAGCGAGGCCCTACAACTACCTGAGGTGCGCTACTTCAAGATCAACAAGTTGGAG 521  
161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180  
162 AAAGCTTGCTGTCAGACACACACCTTCTTCTGCGCAATCCTGAGCACATGGAAATGCAG 581  
181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
182 CAGAACCTAGACTATTACCAACCATGTCTGGAGTGAAGAGGCCGACTTCAAGATCTT 641  
201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGln 220  
202 GAGACTCAACCCCATATGCAAGATTTCCAGTGGAGTGGAGTCTACTCAGAGGAACAG 701  
221 ProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
222 CCACAGAGAGGTGTCGCCACCTAGAGCGCGCTGCAAGAACTATTGTGGCTATGAG 761  
241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
242 GAGTGGCGTGGCTCTCGGAAGGCCCTATGACTACGATGGCTACAACTACCTTGAGTAC 821  
261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
282 AACGCTGACCTCTTCCAGGCCATCACAGATCATTACATCCAGGTCCTCAACTGTGAACAG 881  
281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
282 AACGTGTACGAGAGCTTGTCTCCACCCCAAGTCGAGAGAGCCCTTTGAGACTTCTC 941  
301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
321 CCATCGCATTAATAATTATCTGAGTTTGCCTACTATAACATTTGGGAATTATACACAGCT 1001  
321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
1002 GTTGAATGTGCAGAGACCTACTCTCTCTTCTCCCAATGACGAGGTGATGAACCAAAAT 1061  
341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluHisThrArgSerIleGlyProArgGlu 360  
1062 TTGGCCATTATGTCAGCTATCTTGGAGAGAAACACACACAGATCCATCGGCCCGGTGAG 1121  
361 SerAlaLysGluTyrArgGlnArgSerLeuLeuLysGluLeuPhePheAlaTyr 380  
1122 AGTGCAGAGGAGTACCGACAGCGAAGCCTACTGGAAAAAGAACTGCTTTCTTCGCTTAT 1181  
381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400  
1182 GATGTTTTTGAATTCCTTTTGTGGATCCGATTCATGGACTCCAGGAGAGTCAATCCC 1241  
401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
1242 AAGAGATTGCAAGAGAAACAGAAAGTCAGACGGGAAACAGCGCTACGCTATCCACAGGAG 1301  
421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
1302 ATTGGGAACCTTATGAAGAAATCAGAACCTTTGTGGAGAGAGAGACACAGAGTCACTG 1361  
441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
1362 GATGTGACAGACTGACCCGGAGAGGTGGCCCTGCTGTATGAAGGCATCAGTCTCACC 1421  
461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
1422 ATGAATCCAACTCCTGAATGGTTCCAGCGGGGTGGATGGAGCGGGTAACTCTGAC 1481  
481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
1482 CACGAGTGTACGAGACTGCAGAGACTGACCAATGTGGCAGCAACTCTCAGGAGATGGCTAC 1541

501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
1542 CGGGGTGACAGCTCCCACTACATCTCCCAATGAAAAGTTCTATGGTGTCACTGTCTTCAA 1601  
521 AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
1602 GCCCTCAAGCTGGGCAAGAGCAAGTTCTCTGAGAGTGCCCACTGTACTACAAAC 1661  
541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
1662 GTGACGAGAGGTGGCGGCATCATGGAGTCTACTTCCGCTGGATACGACCCCTCTAC 1721  
561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
1722 TTTTCTCTACTCTCATCTGTGTGGCGCATCTGCATCGAAGAGTCCAGGAGAGAGAG 1781  
581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
1782 GATGATGATCATCTAGCTCCAGCTGGACAACATCATCTCTGAAATGCCGAGACCTCGTGT 1841  
601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
1842 GTCAAGAGCCCGCCAGCTTACACCTTCCGCGACTACAGCGCATCTTTTACCTAAATGG 1901  
621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
1902 GACTTCGATGGCGAAACTTTTATTTCTGAACTGGATGCCAAGACCGTGACGCGAGAG 1961  
641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
1962 GTGACGCTCAGTGTGAAGAGCCGTGGGATTTCTCTTCAAGCACTGAAACCCACATGA 2021  
661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680  
2022 GTGAAGCTGTCAACAGGGGGCAGCGCTGTGCATCGCCCTGTGGTTCACTTCCCTGAGCCCT 2081  
681 ArgHisSerGluArg----- 685  
2082 CGACACAGCGAGCG-GGTGAGAGCAGCTCCAGCGGTGAGAGCGTGTGTGTGTGA 2140  
685 ----- 685  
2141 CCGTTCCCGAGAGCGCCCTTGTGTTGCCCTTCTTCTTCCCAATCCATTCGATGGAGTGGCT 2200  
685 ----- 685  
2201 GAGACACGAAAGGAGCAGCTTGGGACACAGCTCCAGCGCCTGTCAATTATGTTCACATTG 2260  
685 ----- 685  
2261 CCTGTCTCCTCCTGGCCCTGCTGTGAACGGGATCCAGGTGGGAAAGAGGTCAAGACAGG 2320  
685 ----- 685  
2321 GAGCGATGTGAGTTCTTGTTCCTTCCCTTGGCCCCCACTTCAGCTGTCTTTCAGAG 2380  
685 ----- 685  
2381 AGTAGGACCTGTGGCAAGAGGAGATGAGCTGGGGCCATTAAGGAACCTTCTTGTCCCT 2440  
685 ----- 685  
2441 GGGAGTAGCAGCTGAGAGATACCGAGTGTCTGGAGCGGAGGCTCTCTGAATGGCAGG 2500  
686 -----AspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
2501 GGTTTCTCTTGGCAGGACAGGTTGAGGAGATGACCTGGTGAAGATGCTCTTCAGCCCA 2560  
701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
2561 GAAGAGATGTTCTCTCCAGGAGCAGCCCTGGATGCCAGCAGGGCCCCCGGAACT 2620  
721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736



QY 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
Db 1002 GTTGAATGTGCCAAGACCTATCTCTCTTCCCAATACAGAGGTGATGAACCAAAAT 1061  
QY 341 LeuAlaTyrTyrAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
Db 1062 TTGGCTTATTATGAGTATGCTGTGGAGAGAACACACAGATCCATCGCCCCCGGTGAG 1121  
QY 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaTyr 380  
Db 1122 AGTGCCAAGAGTACCCAGACAGGAGGCTACTGAAAAAGAACTGCTTTCTTCGCTTAT 1181  
QY 381 AspValPheGlyIleProPheValAspProAspSerTyrProGluGluValIlePro 400  
Db 1182 GATGTTTTTGGAAATCCCTTTGTGGATCCGGATTCATGGACTCCAGGAGAAGTATCC 1241  
QY 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
Db 1242 AAGAGATTGCAAGAGAAACAGAGTCAAGACGGGAAACACCCCTGATCGCATCTCCAGGAG 1301  
QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluLysThrLysGluSerLeu 440  
Db 1302 ATTGGGAACCTTATGAAGGAAATCCAGACCCCTTGTGGAAGAGAACCAAGAGATCACTG 1361  
QY 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
Db 1362 GATGTGACGACTGACCCGGGAAGGTGGCCCCCTGCTGTATGAGGCATCAGTCTCACC 1421  
QY 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
Db 1422 ATGAACCTCCAACTCTCTGATGGTTCCAGCGGGTGGTGTGAGACGGCGTAATCTCTGAC 1481  
QY 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
Db 1482 CACGAGTGTCCAGGAGCTGCAGAGACTGACCAATGTGCAGCAACCTCAGAGATGGGTAC 1541  
QY 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
Db 1542 CGGGGTACAGCTCCCCACATCTCCCAATGAAAGTTCTATGTTGTCACGTCTTCAAA 1601  
QY 521 AlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
Db 1602 GCCCTCAAGCTGGGCAAGAGGCAAAAGTTCTCTGCAGAGTGGCCACCTGTACTACAAC 1661  
QY 541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
Db 1662 GTGACGGAGAAAGTGGCGCGCATCATGAGTCTTCTCCGCTGGATACGCCCTCTTAC 1721  
QY 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
Db 1722 TTTTCTTACTCTCATCTGGTGTGGCGCACTGCCATCGAAGAGGTCCAGGCAGAGGAAG 1781  
QY 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
Db 1782 GATGATAGTATCCAGTCCACGTGGACAACTGCATCTCGAATGGCGGACCCCTCGTGTGT 1841  
QY 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
Db 1842 GTCAAGAGCCCCAGCCCTACACCTTCGGGACTACAGGCCATCTTTACCTAAATGGG 1901  
QY 621 AspPheAspGlyLysPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
Db 1902 GACTTCGATGGCGGAACTTTTATTTACTGAACCTGGATGCCAAGACCGGTGACGGCAGAG 1961  
QY 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
Db 1962 GTGCAGCCCTCAGTGTGGAAAGACCCGTGGGATTTCTTTCAAGGCATCGAAAAACCATGGA 2021  
QY 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPro 680  
Db 2022 GTGAGGCTGTCCACAGGGGCGAGCGCTGTGCCCTCTGCTTCACTCCCTGAGCCCT 2081  
QY 681 ArgHisSerGluArg----- 685

Db 2082 CGACACAGCGAGCG-GGTGAGAGCAGCTCGAGCGGTGAGAGCAGCTGGTCTGTGTGA 2140  
QY 685 ----- 685  
Db 2141 CCCGTTCCAGAGCGCCCTTGTTTTCCTTCTTCCCAAATCCCATTCOCAGTGGCT 2200  
QY 685 ----- 685  
Db 2201 GAGACACGAAAGAGGACACTTGGGACACCCAGCTCCAAACCCCTGTCTATTATGGTCACTTG 2260  
QY 685 ----- 685  
Db 2261 CTTTGTCTCCCTGGGCTGCTGTGAACGGGATCCAGGTGGGAAAGAGGTCAAGACAGG 2320  
QY 685 ----- 685  
Db 2321 GAGCGATGCTGAGTTCTTGGTTCCCTCCTTGGGCCCCACTTCAGCTGTCTTTTCCAGAG 2380  
QY 685 ----- 685  
Db 2381 AGTAGACCTGCTGGGAAGAGATGAGCTGGGGCCATTAAAGGAACCTTCCTTGTTCCTCT 2440  
QY 685 ----- 685  
Db 2441 GGGAAAGTAGACCTGAGAGATAGCGAGTCTCTGGAGCGGAGCCCTCTCTGAATGGCAGG 2500  
QY 685 -----AspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
Db 2501 GGTTCCTCCTTCAGGACAGGGTGCAGGAGATGACCTGGTGAAGTGTCTTTCAGCCCA 2560  
QY 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
Db 2561 GAAGAGATGCTCTCTCTCCAGGAGCAGCCCTGGATGCCAGCAGGGCCCCCGGAACCT 2620  
QY 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2621 GCACAAAGATGCTCTCTCAGGCAGTGAATCGAAGACCCCAAGATGAGCTA 2668

Search completed: July 18, 2004, 16:08:33  
Job time : 20985 secs

**This Page Blank (uspto)**



```

Percent Similarity: 61.39%      Conservative: 14
Best Local Similarity: 47.52%    Mismatches: 17
Query Match: 5.03%              Indels: 23
DB: 4                            Gaps: 3

US-10-045-815-4 (1-736) x US-09-439-313-271 (1-301)

QY 637 ValThrAlaGluValGln-ProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGl 656
Db 301 GTGACTGCTTCTATAAAACCCAAATGTGGCGCATGATGTTN-TCATCTGGAGGAGA 243
QY 656 uAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPh 676
Db 242 GAACNTCATGGGTGAAGGAGTTCACCAAGGAAAGAGGTGTGCTGTGGCTCTGTGGTT 183
QY 676 eThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMe 696
Db 182 CACCTTGGACCCACTTTATAGAGAAATTGGAGCAATACAGGCTGATGAAGTGAATGCAAT 123
QY 696 tLeuPheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGl 716
Db 122 T-----CTGGATCAAGAA-----CAGCAAGG 102
QY 716 yProProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLe 736
Db 101 GAAGCATGAA-----CTGAATATCAACCCCTAAAGATGAGCT 66
QY 736 u 736
Db 65 A 65

RESULT 3
US-09-232-149A-271/c
; Sequence 271, Application US/09232149A
; Patent No. 6465611
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE
; TITLE OF INVENTION: CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.427C6
; CURRENT APPLICATION NUMBER: US/09/232,149A
; CURRENT FILING DATE: 1999-01-15
; NUMBER OF SEQ ID NOS: 338
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 271
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(301)
; OTHER INFORMATION: n = A,T,C or G
US-09-232-149A-271

Alignment Scores:
Pred. No.: 1,23e-12      Length: 301
Score: 194.50           Matches: 48
Percent Similarity: 61.39%      Conservative: 14
Best Local Similarity: 47.52%    Mismatches: 17
Query Match: 5.03%              Indels: 23
DB: 4                            Gaps: 3

US-10-045-815-4 (1-736) x US-09-232-149A-271 (1-301)

QY 637 ValThrAlaGluValGln-ProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGl 656
Db 301 GTGACTGCTTCTATAAAACCCAAATGTGGCGCATGATGTTN-TCATCTGGAGGAGA 243
QY 656 uAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPh 676
Db 242 GAACNTCATGGGTGAAGGAGTTCACCAAGGAAAGAGGTGTGCTGTGGCTCTGTGGTT 183
QY 676 eThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMe 696
Db 182 CACCTTGGACCCACTTTATAGAGAAATTGGAGCAATACAGGCTGATGAAGTGAATGCAAT 123
QY 696 tLeuPheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGl 716
Db 122 T-----CTGGATCAAGAA-----CAGCAAGG 102
QY 716 yProProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLe 736
Db 101 GAAGCATGAA-----CTGAATATCAACCCCTAAAGATGAGCT 66
QY 736 u 736
Db 65 A 65

RESULT 2
US-09-352-616A-271/c
; Sequence 271, Application US/09352616A
; Patent No. 6395278
; GENERAL INFORMATION:
; APPLICANT: Dillon, Davin C.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang, Yuqui
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF PROSTATE CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.427C8
; CURRENT APPLICATION NUMBER: US/09/352,616A
; CURRENT FILING DATE: 1999-07-13
; NUMBER OF SEQ ID NOS: 472
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 271
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(301)
; OTHER INFORMATION: n = A,T,C or G
US-09-352-616A-271

Alignment Scores:
Pred. No.: 1,23e-12      Length: 301
Score: 194.50           Matches: 48
Percent Similarity: 61.39%      Conservative: 14
Best Local Similarity: 47.52%    Mismatches: 17
Query Match: 5.03%              Indels: 23
DB: 4                            Gaps: 3

US-10-045-815-4 (1-736) x US-09-352-616A-271 (1-301)

QY 637 ValThrAlaGluValGln-ProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGl 656
Db 301 GTGACTGCTTCTATAAAACCCAAATGTGGCGCATGATGTTN-TCATCTGGAGGAGA 243
QY 656 uAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPh 676

```

QY 736 u 736  
 Db 65 A 65  
 RESULT 4  
 US-09-159-812-271/c  
 ; Sequence 271, Application US/09159812A  
 ; Patent No. 6613872  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Xu, Jiangchun  
 ; APPLICANT: Dillon, Davin C.  
 ; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF  
 ; TITLE OF INVENTION: PROSTATE CANCER AND METHODS FOR THEIR USE  
 ; FILE REFERENCE: 210121.428C5  
 ; CURRENT APPLICATION NUMBER: US/09/159,812A  
 ; CURRENT FILING DATE: 1998-09-23  
 ; NUMBER OF SEQ ID NOS: 306  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 271  
 ; LENGTH: 301  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapien  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)....(301)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-159-812-271

Alignment Scores:  
 Pred. No.: 1.23e-12 Length: 301  
 Score: 194.50 Matches: 48  
 Percent Similarity: 61.32% Conservative: 14  
 Best Local Similarity: 47.52% Mismatches: 17  
 Query Match: 5.03% Indels: 23  
 DB: 4 Gaps: 3

US-10-045-815-4 (1-736) x US-09-159-812-271 (1-301)

QY 637 ValThrAlaGluValGln-ProGlnCysGlyArgAlaValGlyPheSerGlyThrGl 656  
 Db 301 GTGACTGCTTCTATAAACCCAAATGTGGCGCATGATCAGTTN-TCACTGGAGGAGA 243  
 QY 656 uAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaLeuAlaLeuTrpPh 676  
 Db 242 GAACNCTCATGGGTCAAGGCAGTCACCAAGGAAAGAGGTGTCTGTGGCTCTGTGGTT 183  
 QY 676 eThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMe 696  
 Db 182 CACCTTGGACCCACTTTATAGAGATTGGAGCGAATACAGCTGATGAAGTATGCAAT 123  
 QY 696 tLeuPheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGl 716  
 Db 122 T-----CTGGATCAAGAA-----CAGCAAGG 102  
 QY 716 yProProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLe 736  
 Db 101 GAAGCATGAA-----CTGAAATATCAACCTTAAGATGAGCT 66

RESULT 5  
 US-09-636-215-271/c  
 ; Sequence 271, Application US/09636215  
 ; Patent No. 6620922  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Xu, Jiangchun  
 ; APPLICANT: Dillon, Davin C.  
 ; APPLICANT: Mitcham, Jennifer L.  
 ; APPLICANT: Harlocker, Susan L.  
 ; APPLICANT: Jiang, Yuqi

; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: Kalos, Michael D.  
 ; APPLICANT: Fanger, Gary R.  
 ; APPLICANT: Retter, Marc W.  
 ; APPLICANT: Stolk, John H.  
 ; APPLICANT: Day, Craig H.  
 ; APPLICANT: Vedwick, Thomas S.  
 ; APPLICANT: Carter, Darrick  
 ; APPLICANT: Li, Samuel  
 ; APPLICANT: Wang, Aijun  
 ; APPLICANT: Skelky, Yasir A.W.  
 ; APPLICANT: Hepler, William  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
 ; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
 ; FILE REFERENCE: 210121.42717C17  
 ; CURRENT APPLICATION NUMBER: US/09/636,215  
 ; CURRENT FILING DATE: 2000-08-10  
 ; NUMBER OF SEQ ID NOS: 852  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 271  
 ; LENGTH: 301  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapien  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)....(301)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-636-215-271

Alignment Scores:  
 Pred. No.: 1.23e-12 Length: 301  
 Score: 194.50 Matches: 48  
 Percent Similarity: 61.32% Conservative: 14  
 Best Local Similarity: 47.52% Mismatches: 17  
 Query Match: 5.03% Indels: 23  
 DB: 4 Gaps: 3

US-10-045-815-4 (1-736) x US-09-636-215-271 (1-301)

QY 637 ValThrAlaGluValGln-ProGlnCysGlyArgAlaValGlyPheSerGlyThrGl 656  
 Db 301 GTGACTGCTTCTATAAACCCAAATGTGGCGCATGATCAGTTN-TCACTGGAGGAGA 243  
 QY 656 uAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaLeuAlaLeuTrpPh 676  
 Db 242 GAACNCTCATGGGTCAAGGCAGTCACCAAGGAAAGAGGTGTCTGTGGCTCTGTGGTT 183  
 QY 676 eThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMe 696  
 Db 182 CACCTTGGACCCACTTTATAGAGATTGGAGCGAATACAGCTGATGAAGTATGCAAT 123  
 QY 696 tLeuPheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGl 716  
 Db 122 T-----CTGGATCAAGAA-----CAGCAAGG 102  
 QY 716 yProProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLe 736  
 Db 101 GAAGCATGAA-----CTGAAATATCAACCTTAAGATGAGCT 66

RESULT 6  
 US-09-685-165A-271/c  
 ; Sequence 271, Application US/09685165A  
 ; Patent No. 6630305  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Xu, Jiangchun  
 ; APPLICANT: Dillon, Davin C.  
 ; APPLICANT: Mitcham, Jennifer L.  
 ; APPLICANT: Harlocker, Susan L.  
 ; APPLICANT: Jiang, Yuqi







```

/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature
/ LOCATION: (765001)..(780000)
/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature
/ LOCATION: (780001)..(795000)
/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature
/ LOCATION: (795001)..(810000)
/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature
/ LOCATION: (810001)..(825000)
/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature
/ LOCATION: (825001)..(840000)
/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature
/ LOCATION: (840001)..(855000)
/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature
/ LOCATION: (855001)..(870000)
/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature
/ LOCATION: (870001)..(885000)
/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature
/ LOCATION: (885001)..(900000)
/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature
/ LOCATION: (900001)..(915000)
/ OTHER INFORMATION: n=a or c or g or t
/ NAME/KEY: misc feature

```

```

Alignment Scores:
Pred. No.: 0.00584 Length: 1230025
Score: 162.50 Matches: 146
Percent Similarity: 33.53% Conservative: 86
Best Local Similarity: 21.10% Mismatches: 257
Query Match: 4.20% Indels: 203
DB: 4 Gaps: 50

```

US-10-045-815-4 (1-736) x US-09-198-452A-1 (1-1230025)

```

Qy 51 TtpProGlyValValLeuSerMetGluArgAlaLeuArgSerArgAlaAlaLeuArgAla 70
Db 1082828 TGGCCCAAGTG-----GAAGTATCCGACGAGGCGATACCTCTTCA 1082784
Qy 71 LeuA:GLeu-----ArgCysArgThrGlnCysAla 80
Db 1082783 CAACGACAAATTGATTTTGGCCCTTATGTAAGTGCTGAAGCCTGCCACACTCTCTTA 1082727
Qy 81 AlaAspPheProTprGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100
Db 1082726 -----GAGTTATCAGCAATGTTTCCT----- 1082703
Qy 101 AlaGlyAlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgArgAlaAlaCys--- 119
Db 1082702 -----CTGAGACTTGTTCAGATCGAGAGTTGCTTAAAGGAAACGCCCTTGCATT 1082652
Qy 120 -----LeuArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMet 136
Db 1082651 CTTTATGACATGAACGGTGCTTCTCTTGTGCGCTATATGCTCCTGGAAGATAT 1082592
Qy 137 GluLeuGluPheArgGlyArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLeu 156
Db 1082591 CAAGCAACTTAGACAAG-----GCGATCCTCTTTCTTAAA 1082556
Qy 157 AsnLysLeuGluLysAlaValAlaAlaHisThrPhePheValGlyAsnProGluHis 176
Db 1082555 GGAAAGATCGAAGAGTGTGCTCAAGGACTTAGAAAAAGTGATTTCAAAAGGCTTCGGATAAC 1082496
Qy 177 MetGluMetGlnGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAsp 196

```

```

Db 1082495 TTAGAATTTGAGCAAGCTGCGAATTACTACCGACGTTCTCCCTAATTAAGCAAGCTATG 1082436
Qy 197 PheLysAspLeuGluThrGlnProHisMetGlnGluPhe----- 209
Db 1082435 GCAAGCAACAGGTGGAAAAATTTCAATTTTCAAAATATTGATGCTCTTGGTCTTTACAGA 1082376
Qy 210 -----ArgLeuGlyVal 213
Db 1082375 CATAAACAACGACCATCTCCTACGCTACTTACTGTCCGCTCAGGAAGCTTCTTGGGCA 1082316
Qy 214 ArgLeuTyrSer---GluGluGlnProGlnGluAlaValProHisLeuGluAlaLeu 232
Db 1082315 CGCAATTTCTTCTTTTGGAAATGCTCAAGAGATCAAGACTTGTCTTCTCTTTTATT 1082256
Qy 233 GlnGluTyrPheValAla-----TyrGluGluCysArgAlaLeuCysGluGlyProTyr 250
Db 1082255 TTGCAATATTATGTTAGCAACCCCTACATTCTTAAAGAAATCTGCACACCTCTTCCCTA 1082196
Qy 251 AspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaAsp-----LeuPheGln 266
Db 1082195 GAAATTCCTACTCTCTCTACGTA---TTGAATGCGGAGTCTCCTCCCGCTACGTTCT 1082139
Qy 267 AlaIleThrAspHisTyrIleGlnValLeu-----AsnCysLysGlnAsn 281
Db 1082138 CCAAAACAGGATATGAAAAGAACTTCTTGATCTAGCCTATCGCAATGCCAAGCGTAT 1082079
Qy 282 CysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu--- 300
Db 1082078 GCAGCAACTACACTTCTCTCATCAGCTCCCTTACCAAGCTTTTAAAGATAATACCTGG 1082019
Qy 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320
Db 1082018 ATGTCATATATCCGTATCGGATAGTATCGTATAGTCT-----CATATGAAGT 1081965
Qy 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340
Db 1081964 GCTCAGCTACTGGAGTGACATTGTTATTTGAAAAATATAGGTTTCGATCCCAAGCAA--- 1081908
Qy 341 LeuAlaTyrTyrAlaAlaMetLeuGluGluHisThrArgSerIleGlyProArgGlu 360
Db 1081907 -----TATCGTACGTTTTCCTA-----GAT 1081887
Qy 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr 380
Db 1081886 TCTGAAAGACTCAGAAATGACATGCTGCTCTATTAGAAAGAGTCTCTTA----- 1081839
Qy 381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400
Db 1081838 -----CGAAGGTTTTCATTACAGCACAGCTCTTCTCGATGATGATTGTG--- 1081794
Qy 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420
Db 1081793 -----GTTGATCGGGGAAACGCCACTATAACAGACAAAAAATAATCCAACT 1081743
Qy 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSer--- 439
Db 1081742 CTCAATCTTACAGGATTCAGGTTGTTACTATCGAAAAGAAAAAGTAATCATAGTCGA 1081683
Qy 440 ---LeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSer 458
Db 1081682 GGGTTAAATAGGAGNAGATTTTCTCGAGAGCTTCCCA-----GAGGTTTTTCT 1081632
Qy 459 LeuThrMetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIle 478
Db 1081631 CTTCCCGCACCTCTAAATTTATTACAAATTTTTCAG-----ATCCTT 1081590
Qy 479 SerAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAsp 498
Db 1081589 CGTGAT-----GAG 1081581
Qy 499 GlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrVal 518
Db 1081580 GCACATCGCTTTGCGATTAGTAAAGCAC-----AGAAAAAAGAGAAAGGCTTTA 1081530

```

QY	519	PhelysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyr	538	Percent Similarity: 34.89%	Conservative: 54		
Db	1081529	TTTGAACAGGAAAAATCCCTGGATCGGAGAGTA	1081494	Best Local Similarity: 22.25%	Mismatches: 160		
QY	539	TyrAsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrPro	558	Query Match: 3.17%	Indels: 118		
Db	1081493	-----AAGCGAAAAAGATTGTCAGAAA	1081470	DB: 4	Gaps: 20		
QY	559	LeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGlu	578	US-10-045-815-4 (1-736) x US-09-738-946-5 (1-2595)			
Db	1081469	-----TTTAAAGTTGGAAACAGGTATATGCTTAGCTCCCGAAGAGCTGGAAGCT---	1081419	QY	355	SerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLys-Gl	374
QY	579	ArgLysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeu	598	Db	782	AGCTGTGGCATCTGCCAATCTGGAATGAGCGCGCGCGCCAACTGCTCCACGCCCA	841
Db	1081419	-----	1081419	QY	374	uLeuLeuPhePheAlaTyrAspValPheGlyIleProPheValAspProAspSerTyrTh	394
QY	599	ValCysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeu	618	Db	842	GATGATG-----ACGCCTGCCACCCAGGGCGAGATTGG--	875
Db	1081418	-----ATTCTGGGTAACTAAGAAAGATATTGCAGTCTTTCGAAGG	1081374	QY	394	rProGluGluVal-----IleProLysArgLeuGlnGlu	406
QY	619	AsnGlyAspPheAspGlyGlyAsn---PheTyrPheThrGluLeuAspAlaLysThrVal	637	Db	876	-----CCCCCGAGGTAGACAACCTGCTGAATCTCTCGGCGAGCGCGCAGGAGAAGGC	934
Db	1081373	CAAAAAGACTTCAACAAATCTGATTAACTCTACCTTCAGATTCCTCAAGTAAGAAGTCT	1081314	QY	406	sGlnLysSerGluArgGluThrAlaValArgIleSerGln-----	419
QY	638	ThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsn	657	Db	935	TCGAGCTCGGAAACGGAACAGGTCAAAAGCAGCAGCAGCAGCAGCATCATCACCATAA	994
Db	1081313	TCAGGTGAA-----GGAAGA-----TCTTGAACAGAGACT	1081284	QY	420	-----GluIleGlyAsnLeuMetLysGluIleGluThrLeuVa	432
QY	658	ProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThr	677	Db	995	TGCGAAAAAGAGCTCCAGCTATCAATCGTCTGGCGATGCCAGCTTCATGGGATCAGG	1054
Db	1081283	TCCTCTTCAATTTCTCATCAATACGATCTTTGGTTGTGTCAACCTTTAATGATCAGC	1081224	QY	432	IleGluLysThrLysGluSerLeu-----AspValSerArgLeuThrArgGluGl	449
QY	678	Leu-----AspProArgHisSerGlu	684	Db	1055	AAAGTGGCGCCGTATGAGGATCTGTCCCGTAACATCATCAGCGACATGAACCATATGG	1114
Db	1081223	TGGACAGCGGCCACATGAAGTCTCTGTATATCTGAA	1081188	QY	449	YGlyProLeuLeuTyrGluGlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTy	469
RESULT 9				Db	1115	TCGTCCGTGGTGGACAGCTTCTCTGGGATGGAGACGGGCGCTGAAGATCCTCAACGAGGT	1174
US-09-738-946-5				QY	469	rGlnArgValValMetAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLe	489
Sequence 5, Application US/09738946				Db	1175	TGCAAGCATGTACAACGCGAGGAGCCTTCAGGATGCG-----CAAAGTGT	1219
Patent No. 6579701				QY	489	uThrAsn-----ValAlaAlaThrSerGlyAspGlyTyrArgGlyGln----	503
GENERAL INFORMATION:				Db	1220	GACCAACAGAGCGCCCGATGCACCCGCGTGGCGGTGACAAGATCCGAGGCGATAAGAT	1279
APPLICANT: EXELIXIS, INC.				QY	504	-----ThrSerProHisThrProAsnGluLysPhe-----Ty	514
TITLE OF INVENTION: DROSOPHILA HOMOLOGUES OF GENES AND PROTEINS IMPLICATED IN CANCER				Db	1280	CAAGTGGGTGGTGGCAATGAGCCGGCTGCAGCAATGCTGTGGTATCTGACCAATCAGAT	1339
TITLE OF INVENTION: METHODS OF USE				QY	514	rGlyValThrValPheLysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSe	534
FILE REFERENCE: EX00-043C				Db	1340	TGACTCTGTGTGTATCGTGTTAACACGATGAAGATAATGGCATCTTTGGSC-----	1391
CURRENT APPLICATION NUMBER: US/09/738,946				QY	534	rAlaHisLeuTyrTyrAsnValThrGluLysValArgArgIleMetGluSerTyrPheAr	554
CURRENT FILING DATE: 2000-12-14				Db	1392	-----AACTACCATCAGGAGCGGCGAGGCGCAATGGTCGTTGTAT-----	1436
PRIOR APPLICATION NUMBER: 60/170,832				QY	554	gLeuAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGl	574
PRIOR FILING DATE: 1999-12-14				Db	1437	-----CCGGATCGGGAACCTCACTACGTC-----	1460
PRIOR APPLICATION NUMBER: 60/170,838				QY	574	uValGlnAlaGluArgLysAspSerHisProValHisValAspAsnCysIleLeuAs	594
PRIOR FILING DATE: 1999-12-14				Db	1461	-----ATGCATGTGACAAT-----	1475
PRIOR APPLICATION NUMBER: 60/178,580				QY	594	nAlaGluThrLeuValCysValLysGluProProAlaTyrThrPheArgAspTyrSerAl	614
PRIOR FILING DATE: 2000-01-28				Db	1476	-----CCCCAAAGGATGGCGCGGTTATAACGCGC	1504
PRIOR APPLICATION NUMBER: 60/185,879				QY	614	alleLeuTyrLeuAsnGlyAspPheAsp-----GlyGlyAsnPheTyrPheTh	630
PRIOR FILING DATE: 2000-02-29				Db	1505	CATATACTACCTGAATATCACTACCTGCGGGAAGTGGCGGCAATTTCTGCGAATTCG	1564
PRIOR APPLICATION NUMBER: 60/185,880				QY	630	rGluLeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGl	650
PRIOR FILING DATE: 2000-02-29				Db			
PRIOR APPLICATION NUMBER: 60/186,150							
PRIOR FILING DATE: 2000-03-01							
PRIOR APPLICATION NUMBER: 60/189,701							
PRIOR FILING DATE: 2000-03-15							
NUMBER OF SEQ ID NOS: 14							
SOFTWARE: PatentIn version 3.0							
SEQ ID NO 5							
LENGTH: 2595							
TYPE: DNA							
ORGANISM: Drosophila melanogaster							
US-09-738-946-5							
Alignment Scores:							
Pred. No.:	0.00492						
Score:	122.50						



[illegible]

Db 24326 TTACTTATTTCTATGCTACACTGGCAGACGCGCATGGCGTTTGGCAGCAGTCTTGGGA 24385  
Qy 121 GARGCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeu----- 138  
Db 24386 GGAGTGC-----AACCTCAAGGAGCTGCAGAACTGCTTAAGCA 24424  
Qy 139 -----GluPheArgLysArgSerProTyrsnTyrluGlnVa 151  
Db 24425 AAACCTTGAGGACCTATGAGCGGCTTCAACGAGCGCTCC----- 24464  
Qy 151 lAlaTyrlPheLysIleAsnLysLeuGluLysAlaValAlaAlaHisThrPhePheVa 171  
Db 24465 -----GTGCCCGGACCTGGCGGACAT 24487  
Qy 171 lGlyAsnProGluHisMet-----GluMetGlnGlnAsnLeuAspTyrlGlnThrMe 189  
Db 24488 CATTTTCCCGAAGCGCTGCTTAAACCCCTGCAACAGGGTCTG----- 24530  
Qy 189 tSeiGlyValLysGluAlaAspPheLysAspLeuGluThrGlnProHisMetGlnGluPh 209  
Db 24531 -----CCAGACTTCACCGTCAAGCATGTTCGAACTT 24565  
Qy 209 eArgLeuGlyValArgLeuTyrsrGluGluGlnProGlnGluAlaValProHisLeuGl 229  
Db 24566 TAGGAACCTTATCTCTA-----GAGCGCTCAGGAATCTTGCCCGCCACCTG 24610  
Qy 229 uAlaAlaLeuGlnGluTyrlPheVal-----AlaTyrlGluGluCysArgAlaLeuCysGl 247  
Db 24611 CTGTGCACCTTCTACGCACTTTGTGCCCATTAAGTACCGCAATGCCCTCGCGCGCTTG 24670  
Qy 247 uGlyPro--TyrsPtyrAspGlyTyrsnTyrluGluTyrsnAlaAspLeuPheGl 266  
Db 24671 GGGCCACTGCTACCTTCGACGTAGCACTACCTTCCCTACCACTGCACATTAAGGA 24730  
Qy 266 nAlaIleThrAspHisTyrlleGlnValLeuAsnCysLysGlnAsnCysValThr----- 284  
Db 24731 AGACGTGAGCGGTGACGGTCTACTGGAGTGTCACTGTGCTGCAACCTATGCACCCGCA 24790  
Qy 285 -----G1 285  
Db 24791 CGGCTCCTCGTTTGCATTCGACGTCTTAACGAAAGTCAATATTCGTACTCTTGA 24850  
Qy 285 uLeuAlaSerHisProSerArgGluLysPro----- 295  
Db 24851 GCTGAGGCTCCCTCGCTGACGAAAGTCCGCGCTCCGGGTGAACTCACTCCGGG 24910  
Qy 296 -----PheGluAspPheLeuProSerHisTyrsnTyrluGlnPh 309  
Db 24911 GCTGTGGACGTGCGTTTACCTTGGCAATTTGTACCTGAGCACTACCACGCCACGAGAT 24970  
Qy 309 eAlaTyrlTyrsnIleGlyAsnTyrlGlnAlaGlyLysAlaLysThrTyrluLe 329  
Db 24971 TAGGTTCTAC----- 24980  
Qy 329 uPhePheProAsnAspGluValMetAsnGlnAsnLeuAlaTyrlTyrlAlaAlaMetLeuGl 349  
Db 24981 -----GAAGACCAATCCCGCCCGCCCAATGCGAGCTTACCAGCTCGGCTTAC 25030  
Qy 349 yGluGluHisThrArgSerIleGlyProArgGluSerAlaLysGluTyrlArgGlnArgse 369  
Db 25031 CCAGGGCCACATT-----CTTGCCCAATTGCAAGCCATCAACAGCCCGCAAGATT 25084  
Qy 369 rLeuLeuGluLysGluLeuLeuPheAlaTyrlAspValPheGlyIleProPheValas 389  
Db 25085 TCTGTACGAAAGGACCG-----GGGGT---TACTTGGGA 25117  
Qy 389 pProAspSerTrpThrProGluGluValIleProLysArgLeuGlnGluLysGlnLysse 409  
Db 25118 CCCCAGTCCGCGAGGAGCTCAACCCATCCCGCCCGCCGAGCCCTATCAGCAGCA 25177  
Qy 409 rGluArgGluThrAla 414  
Db 25178 GCCCGGGCCCTTGCT 25193

RESULT 13  
US-08-735-609-4  
; Sequence 4, Application US/08735609  
; Patent No. 5955360  
; GENERAL INFORMATION:  
; APPLICANT: Chamberlain, Jeffrey S.  
; APPLICANT: Amalfitano, Andrea  
; APPLICANT: Hauser, Michael A.  
; APPLICANT: Kumar-Singh, Rajendra  
; APPLICANT: Hartigan-O'Connor, Dennis J.  
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States Of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/735,609  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: UM-02484  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 34303 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; DESCRIPTION: /desc = "DNA"  
US-08-735-609-4  
Alignment Scores:  
Pred. No.: 2,34 Length: 34303  
Score: 115.50 Matches: 96  
Percent Similarity: 30.94% Conservative: 42  
Best Local Similarity: 21.52% Mismatches: 146  
Query Match: 2.98% Indels: 162  
DB: 2 Gaps: 20  
US-10-045-815-4 (1-736) x US-08-735-609-4 (1-34303)  
Qy 42 GlyThrAlaAlaTyrlAlaArgGlyAspTrpProGlyValValLeuSerMetGluArgAla 61  
Db 26620 GGAGATGCGACGCAAGCTAGAGAAAC-----ATTGCACTACACTTTTCGACAGG 26670  
Qy 62 LeuArgSer-----ArgAlaAlaLeuArgAlaLeuArgLeuArg 74  
Db 26671 CTACGTACCCAGGCGCTGCAAGATCTCCACGTGAGCTCTGCACTGGTCTCTCTA--- 26727  
Qy 75 CysArgThrGlnCysAlaAlaAspPheTrpGluLeu-----AspProAspTrpSer 92  
Db 26728 -----CCTTGAATTTTGCACGAAACCGCTTTGGGCA 26760  
Qy 93 ProSerPro-----AlaGlnAlaSerGlyAlaGlyAlaLeuArgAsp----- 106  
Db 26761 AAACGTGCTTCATTCACGCTCAAGGCGAGCGCGCGGCTACGTCCGCGACTGGGT 26820  
Qy 107 -----LeuSerPhePheGlyGlyLeuLeuArgArgAlaAla-CysLeuAr 121

Db 26821 TTACTTATTCTATGCTACACCTGCAGACGGCCCTTTGGCAGCAGTCTTGA 26880  
QY 121 GARGCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeu----- 138  
Db 26881 GGAGTGC-----AACTCAAGAGCTGCAGAAACTGCTAAAGCA 26919  
QY 139 -----GluPheArgLysArgSerProTyrAsnTyrLeuGlnVa 151  
Db 26920 AAACCTGAAGACCTATGACGGCTTCAACGAGCGCTCC----- 26959  
QY 151 lAlaTyrPheLysLeuLeuLysLeuGluLysAlaValAlaAlaHisThrPhePheVa 171  
Db 26960 -----GTGCCCGCGCCTTCAACGAGCGCTCC----- 26982  
QY 171 lGlyAsnProGluHisMet-----GluMetGlnGlnAsnLeuAspTyrTyrGlnThrMe 189  
Db 26983 CATTTTCCCGACGCGCTGCTTAAACCCCTGCACAGGCTCG----- 27025  
QY 189 tSerGlyValLysGluAlaAspPheLysAspLeuGluThrGlnProHisMetGlnGluPh 209  
Db 27026 -----CCAGACTTCACCATGCAAGCATGTTGCAGAACTT 27060  
QY 209 eArgLeuGlyValArgLeuTyrSerGluGluGlnProGlnGluAlaValProHisLeuG1 229  
Db 27061 TAGGAACCTTATCTCTA-----GAGCGCTCAGAACTTGTGCCGCCACCTG 27105  
QY 229 uAlaAlaLeuGlnGluTyrPheVal-----AlaTyrGluGluCysArgAlaLeuCysG1 247  
Db 27106 CTGTGCACCTCTAGCGACTTGTGCCATTAAAGTACCGCAATGCCCTCGCGCTTTG 27165  
QY 247 uGlyPro-----TyrAspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaAspLeuPheG1 266  
Db 27166 GGGCCACTGTACCTTCTGCAGCTAGCAACTCTGCTTGCCTACCACTGTGACATAATGA 27225  
QY 266 nAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGlnAsnCysValThr----- 284  
Db 27226 AGACGTGAGCGGTGACGGTCTACTGGAGTGTACCTGCTGCTGCACTATGCACCCGCA 27285  
QY 285 -----CGCTCCCTGTTTGCATTCGACGCTTACAGAAAGTCAAAATTATCGTACCTTTGA 27345  
Db 27286 CCGCTCCCTGTTTGCATTCGACGCTTACAGAAAGTCAAAATTATCGTACCTTTGA 27345  
QY 285 uLeuAlaSerHisProSerArgGluLysPro----- 295  
Db 27346 CCGTGCAGGCTCCCTGCTGACGAAAGTCCGCGGCTCCGGGGTTGAACACTCACTCCGG 27405  
QY 296 -----PheGluAspPheLeuProSerHisTyrAsnTyrLeuGlnPh 309  
Db 27406 GCTGTGACGTCGGCTTACCTTCGCAAAATTGTACTGAGGACTACCACGCCACGAGAT 27465  
QY 309 eAlaTyrTyrAsnIleGlyAsnTyrThrGlnAlaGlyGluCysAlaLysThrTyrLeuLe 329  
Db 27466 TAGGTTCTAC----- 27475  
QY 329 uPhePheProAsnAspGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuG1 349  
Db 27476 -----GAAGACCAATCCCGCCCGCAATGCGGAGCTTACCGCTTGGCTTAC 27525  
QY 349 yGluGluHisThrArgSerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSe 369  
Db 27526 CCAGGGCCACAT-----CTTGCCCAATTGCAAGCCATCAACAAAGCCCGCCAGAGTT 27579  
QY 369 rLeuLeuGluLysGluLeuLeuPheAlaTyrAspValPheGlyIleProPheValAs 369  
Db 27580 TCTGCTCAAGAGGACGG-----GGGTTT-----TACTTGA 27612  
QY 389 pProAspSerTrpThrProGluGluValIleProLysArgLeuGlnGluLysLysSe 409  
Db 27613 CCCCAGTCCCGCGAGGAGTCAACCAATCCCGCCCGCGCAGCCCTATCAGCAGCA 27672  
QY 409 rGluArgGluThrAla 414

Db 27673 GCGCGGGCGCCTTGCT 27688  
RESULT 14  
US-08-735-609-4  
; Sequence 4, Application US/08735609  
; Patent No. 5994132  
; GENERAL INFORMATION:  
; APPLICANT: Chamberlain, Jeffrey S.  
; Amalfitano, Andrea  
; Hauser, Michael A.  
; Kumar-Singh, Rajendra  
; Hartigan-O'Connor, Dennis J.  
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States Of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/735,609  
; FILING DATE: 23-Oct-1996  
; CLASSIFICATION: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: UM-02484  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 34303 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLSCULE TYPE: other nucleic acid  
; DESCRIPTION: /desc = "DNA"  
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-08-735-609-4  
Alignment Scores:  
Pred. No.: 2,34 Length: 34303  
Score: 115.50 Matches: 96  
Percent Similarity: 30.94% Conservative: 42  
Best Local Similarity: 21.52% Mismatches: 146  
Query Match: 2.98% Indels: 162  
Gaps: 20  
US-10-045-815-4 (1-736) x US-08-735-609-4 (1-34303)  
QY 42 GlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArgAla 61  
Db 26620 GGAGATGACGCGCAAGCTAGAGAAAC-----ATTGCACTACACCTTTTCGACAGG 26670  
QY 62 LeuArgSer-----ArgAlaAlaLeuArgAlaLeuArgLeuArg 74  
Db 26671 CTACGTACCCAGGCGCTGCAAGATCCNACCTGAGGCTCTGCAACCTGGTCTCTTA--- 26727  
QY 75 CysArgThrGlnCysAlaAlaAspPheProTrpGluLeu-----AspProAspTrpSer 92  
Db 26728 -----CCTTGGAAATTTGCAGAAACCGCTTGGGCA 26760  
QY 93 ProSerPro-----AlaGlnAlaSerGlyAlaGlyAlaLeuArgAsp----- 106  
Db 26761 AAACGTGCTTCTATCCACGCTCAAGGGGAGCGCGCGGACTACGTCGCGGACTGCGT 26820





Qy 93 ProSerPro-----AlaGlnAlaSerGlyAlaGlyAlaLeuArgAsp----- 106  
Db 26761 AAACGTGCTTCAATCCACGCTCAAGCGCGAGCGCGCGACTACGTCGCGACGTGCGT 26820  
Qy 107 -----LeuSerPhePheGlyGlyLeuLeuArgAlaAla-CysLeuAr 121  
Db 26821 TTACTTATTCTATGCTACACCTGGCAGCGCCATGGCGTTGGCAGCAGTGTGGA 26880  
Qy 121 gArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeu----- 138  
Db 26881 GGAGTGC-----AACTCAAGAGAGTGCAGAACTGTCTAAAGCA 26919  
Qy 139 -----GluPheArgLysArgSerProTyrAsnTyrLeuGlnVa 151  
Db 26920 AAACCTGAAGACCTATGGAGCGCTTCAAGAGCGCTCC----- 26959  
Qy 151 lAlaTyrPheLysIleAsnLysLeuGluLysAlaValAlaAlaAlaHisThrPhePheVa 171  
Db 26960 -----GTGGCGCGCACCTGGCGGACAT 26982  
Qy 171 lGlyAsnProGluHisMet-----GluMetGlnGlnAsnLeuAspTyrTyrGlnThrMe 189  
Db 26983 CATTTTCCCGAAGCGCTGTAAACCCCTGCAACAGGGTCTG----- 27025  
Qy 189 tSerGlyValLysGluAlaAspPheLysAspLeuGluThrGlnProHisMetGlnGluPh 209  
Db 27026 -----CCAGACTTCACCGAGTCAAGACATGTTGCAGAACTT 27060  
Qy 209 eArgLeuGlyValArgLeuTyrSerGluGluGlnProGlnGluAlaValProHisLeuGl 229  
Db 27061 TAGGAACCTTATCTCA-----GAGCGCTCAGGAATCTTGGCGGCCACTG 27105  
Qy 229 uAlaAlaLeuGlnGluTyrPheVal-----AlaTyrGluGluCysArgAlaLeuCysGl 247  
Db 27106 CTGTGCACTTCTAGCAGCTTTGTGCCCATTAAGTACGGGAAATGCCCTCGCGCGCTTG 27165  
Qy 247 uGlyPro-----TyrAspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGl 266  
Db 27166 GGGCCACTGTACTCTTCGACGTAGCCAACTACCTTGCCTTACCACCTGCACATAATGA 27225  
Qy 266 nAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGlnAsnCysValThr----- 284  
Db 27226 AGACGTGACGGGTACGGTCTACTGGAGTGTCACTGTGCTGCAACCTATGCACCCCGCA 27285  
Qy 285 -----G1 285  
Db 27286 CCGCTCCCTGTTTCAATTCCAGCTGCTTAACGAAAGTCAAATTAATCGGTACTTTGA 27345  
Qy 285 uLeuAlaSerHisProSerArgGluLysPro----- 295  
Db 27346 GCTGCGGGTCCCTGCTGACGAAAGTCCGCGGCTCCGGGGTTGAAACTCACTCCGGG 27405  
Qy 296 -----PheGluAspPheLeuProSerHisTyrAsnTyrLeuGlnPh 309  
Db 27406 GCTGTGGACGTGGCTTACCTTCGCAATTTGTACTCAGGACTACCACGCCACGAGAT 27465  
Qy 309 eAlaTyrTyrAsnIleGlyAsnTyrThrGlnAlaGlyCysAlaLysThrTyrLeuLe 329  
Db 27466 TAGGTTCTAC----- 27475  
Qy 329 uPhePheProAsnAspGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGl 349  
Db 27476 -----GAGACCAATCCGCGCCGCGCAATTCGGAGCTTACCGCTCGGTCAATTAC 27525  
Qy 349 yGluGluHisThrArgSerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSe 369  
Db 27526 CCAGGGCCACATT-----CTTGGCAATTCGAGCCATCAACAAAGCCGCCAAGAGTT 27579  
Qy 369 rLeuLeuGluLysGluLeuLeuPhePheAlaTyrAspValPheGlyIleProPheValAs 389  
Db 27580 TCTGCTACGAAAGGACCG-----GGGGT-----TACTTGA 27612

Search completed: July 18, 2004, 18:12:21  
Job time : 947 secs

Qy 389 pProAspSerTThrProGluValIleProLysArgLeuGlnGluLysGlnLysSe 409  
Db 27613 CCCCAGTCCGCGGAGGAGCTCAACCCATCCCCCGCGCGAGCCCTATCAGCAGCA 27672  
Qy 409 rGluArgGluThrAla 414  
Db 27673 GCGCGGGGCCCTGT 27688

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 10:20:10 ; Search time 977 Seconds  
(without alignments)  
3200.280 Million cell updates/sec

Title: US-10-045-815-4

Perfect score: 736

Sequence: 1 MAVRALKLLTLLAVAAAS.....PPEPAQESLSGSSKPKDEL 736

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373663 seqs, 2124099041 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6745146

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xip  
-Q=/cgn2\_1/USPTO.spool\_p/US10045815/runat\_14072004\_123118\_17231/app\_query.fasta\_1.903  
-DB=N Geneseq\_29Jan04 -OMT=fastap -SUFFIX=oli.rng -MINMATCH=0.1 -LOOPCL=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MAFIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10045815.@CIGN\_1\_1\_885@runat\_14072004\_123118\_17231 -NCPU=6 -ICPU=3  
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_29Jan04:.\*  
1: Geneseqn1980s:.\*  
2: Geneseqn1990s:.\*  
3: Geneseqn2000s:.\*  
4: Geneseqn2001as:.\*  
5: Geneseqn2001bs:.\*  
6: Geneseqn2002s:.\*  
7: Geneseqn2003as:.\*  
8: Geneseqn2003bs:.\*  
9: Geneseqn2003cs:.\*  
10: Geneseqn2004s:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	736	100.0	2600	3	AAC64725 Human tum
2	634	86.1	2242	6	ABZ11346 Human pol
3	433	58.8	2211	6	AAS17573 DNA encod
4	433	58.8	2563	5	Aaf93800 Human CDN
5	433	58.8	2585	4	Aah14589 Human CDN
6	407	55.3	2524	4	Aah14513 Human CDN
7	407	55.3	2577	4	Aad08489 Human sec
8	390	53.0	2615	4	AAD08507 Human sec

9	380	51.6	2753	6	ABK94910	Abk94910 Human nov
10	380	51.6	2753	6	ABK94933	Abk94933 Human nov
11	360	48.9	2829	3	AAC64724	AAC64724 Human tum
12	295	40.1	2993	4	AAH16288	Aah16288 Human CDN
13	275	37.4	2127	6	AAS17572	Aas17572 DNA encod
14	244	33.2	2152	6	ABZ11345	Abz11345 Human pol
15	219	29.8	1309	6	ABZ11344	Abz11344 Human pol
16	188	25.5	1607	6	ABQ76716	Abq76716 Human leu
17	179	24.3	550	7	ACA57060	ACA57060 Human adi
18	172	23.4	1595	7	ACA76943	ACa76943 Human ORF
19	103	14.0	682	4	AAI96734	Aai96734 Human neu
20	101	13.7	652	5	AAF93998	Aaf93998 Primer sp
21	101	13.7	738	4	AAH06612	Aah06612 Human CDN
22	91	12.4	2322	3	AAH06885	Aah06885 Human CDN
23	91	12.4	2322	3	AAC64727	AAC64727 Mouse tum
24	91	12.4	2416	3	AAC64726	AAC64726 Mouse tum
25	87	11.8	465	8	ACH42115	ACH42115 Human foe
26	80	10.9	703	4	AAH06792	Aah06792 Human CDN
27	73	9.9	711	4	AAI96735	Aai96735 Human neu
C 28	58	7.9	569	4	AAH11909	Aah11909 Human CDN
C 27	56	7.6	681	5	AAS80583	Aas80583 DNA encod
C 30	42	5.7	554	4	AAH11825	Aah11825 Human sec
C 31	37	5.0	162	3	AAC28769	Aac28769 Human sec
C 32	28	3.8	535	5	AAF94143	Aaf94143 Primer sp
33	26	3.5	504	8	ACH28054	ACH28054 Human adu
34	18	2.4	462	8	ACH23862	ACH23862 Human adu
35	17	2.3	60	6	ABN40277	Abn40277 Human spl
36	14	1.9	416	3	AAH30421	Aah30421 Human col
37	13	1.8	554	9	ACF79597	ACf79597 Human pro
C 38	13	1.8	649	9	ACF79593	ACf79593 Human pro
C 39	13	1.8	684	9	ACF79589	ACf79589 Human pro
C 40	13	1.8	685	9	ACF79590	ACf79590 Human pro
41	13	1.8	793	9	ACF79586	ACf79586 Human pro
42	13	1.8	849	8	ADB83188	ADB83188 Human CDN
43	13	1.8	899	9	ACF79594	ACf79594 Human pro
44	13	1.8	940	9	ACF79595	ACf79595 Human pro
45	13	1.8	1023	4	AAS27088	Aas27088 cDNA enco

ALIGNMENTS

RESULT 1

AAC64725  
ID AAC64725 standard; cDNA; 2600 BP.

XX AAC64725;

AC AAC64725;

XX 27-FEB-2001 (first entry)

XX Human tumour suppressor Gros1-S encoding cDNA SEQ ID NO:3.

XX Tumour suppressor; Gros1-L; Gros1-S; cell proliferation; regulation;

XX cancer; cytostatic; gene therapy; ss.

XX Homo sapiens.

XX WO200065047-A1.

XX 02-NOV-2000.

XX 26-APR-2000; 2000WO-JP002731.

XX 26-APR-1999; 99JP-00118806.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX Wadhwa R, Sugihara T, Yoshida A;

XX WPI; 2000-687340/67.

XX P-PSDB; AAB36392.

XX Full-length tumor suppressor genes encoding Gros1-L, Gros1-S and mouse

XX homologs participating in regulation of cell proliferation, useful in

PT development of preventives and remedies of cancer.

XX Claim 1; Page 70-77; 114pp; Japanese.

XX The present sequence encodes the human tumour suppressor designated Gros1  
CC -S. Gros1-L and Gros1-S have cytotostatic activity and can be used in gene  
CC therapy. Gros1-L and Gros1-S genes are useful in the development of drugs  
CC used to treat and prevent cancer

XX SQ Sequence 2600 BP; 608 A; 757 C; 722 G; 513 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 0 Length: 2600  
Score: 736.00 Matches: 736  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 3 Gaps: 0

US-10-045-815-4 (1-736) x AAC64725 (1-2600)

QY 1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaSer 20  
DB 52 ATGGCGGTACGGCGTGTGAAGCTGTGACCACTGCTGGGTGCTGGCGCGCTCC 111  
QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40  
DB 112 CNAAGCGAGTTCAGTTCGAGCGAGGATGGGCGATGTTGACGCTGATCTGCTTCC 171  
QY 41 GluGlyThrAlaAlaValAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
DB 172 GAGGGGACCGGACCTACGCGCGGGGACTGGCGGGGTGGTCTTGAGCATGGAAACGG 231  
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgCysArgThrGlnCysAla 80  
DB 232 GCGCTGCGCTCCGGGCGAGCCCTCGCGCCCTTCGCTGCGCTCGCGACCCAGTGGCC 291  
QY 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
DB 292 GCCGACTTCCCGTGGAGCTGGACCCCGGACTGGTCCCGACCGCGCCAGGCGCTCGGC 351  
QY 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaCysLeu 120  
DB 352 GCGCGCGCTTCGCGGACCTAGCTTCTCGGGGGCCCTTCGCGTCCGCTGCGCTGCGCTG 411  
QY 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe 140  
DB 412 CGCGCTGCTCGGGCGCGCGCGCCCTCGCTCAGCGAGAGATGGAGTTC 471  
QY 141 ArgLysArgSerProTrpAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
DB 472 CGCAAGCGGAGCCCTACCACTACTCGAGTTCGCTGCGCTTCAAGATCAACAAGTTGGAG 531  
QY 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180  
DB 532 AAAGCTGTGTGAGCAGCACACCTTCTCGTGGGCAATCTCGAGCACATGGAAATCGAG 591  
QY 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
DB 592 CAGAACTTAGACTATTACCAACCACTGTCTGGAGTGAAGGAGGCGGACTTCAAGGATCTT 651  
QY 201 GlnThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGln 220  
DB 652 GAGACTCAACCCCATATGCAAGAAATTCGACTGGAGTGGACCTCTACTCAGAGGAACAG 711  
QY 221 ProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
DB 712 CCACAGGAAGCTGTGCGCCACCTAGAGCGCGCTGCAAGAATACTTTGTGGCGCTATGAG 771  
QY 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
DB 772 GAGTGGCGTGGCTCTCGAAGGGCCCTATGACTACGATGGCTACAACTACCTTGAGTAC 831

QY 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
DB 832 AACGCTGACCTTCCAGGCCCATCACAGATCATATACATCCAGTCTCTCAACTGTAGCAG 891  
QY 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
DB 892 AACTGTGTACGGAGCTTCTTCCCAAGTCGAGAGAGCCCTTTGAGAGCTTCTCTC 951  
QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
DB 952 CCATCGCATTAATAATCTGCGATTTGGCTACTATAACATTTGGGAATTATACAAAGCT 1011  
QY 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
DB 1012 GGTGAATGTGCCAAGACCTATCTTCTTCTTCCCAATGACGAGTGTATGAACCAAAAT 1071  
QY 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
DB 1072 TTGGCTATTATGACGCTATGCTTGAGAGAAACACACACAGATCCATCGGCCCGCTGAG 1131  
QY 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr 380  
DB 1132 AGTGCCCAAGGAGTACCGACAGCAAGCTTCTGGAAGAAAGAACTGTTTCTTCGCTTAT 1191  
QY 381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400  
DB 1192 GATGTTTTTGGAAATCCCTTCTGGATCGGATTCATGACCTCCAGAGAAAGTGTATCC 1251  
QY 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
DB 1252 AAGAGATTGCAAGAGAAAACAGAAAGTCAGAACGGGAAACAGCGTACGATCTCCAGGAG 1311  
QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluLysThrLysLysSerLeu 440  
DB 1312 ATTGGAACTTATGAGGAATCGAGACCTTTTGGAGAGAGAACCAAGAGTCACTG 1371  
QY 441 AspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
DB 1372 GATGTGAGCAGACTACCGGGAAGTGGCCCTCTGTATGAAGGCATCAGTCTCACC 1431  
QY 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspCysValIleSerAsp 480  
DB 1432 ATGAATCCAAACTCCTGAATGTTTACAGCGGGTGGTATGGAGCGGCTAATCTCTGAC 1491  
QY 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
DB 1492 CACGAGTGTGAGAGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGTAC 1551  
QY 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
DB 1552 CGGGGTGACAGCTCCCGACATCTCCCAATGAAAGTTCTATGGTGTCTACTGTCTCAAA 1611  
QY 521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
DB 1612 GCCCTCAAGCTGGGCAAGAGGCAAGTTCCTCTGAGAGTGGCCACCTGTACTACAC 1671  
QY 541 ValThrGluLysValArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
DB 1672 GTGACGGAGAAAGTGGCGCATCATCTACTTCCGCTGGATACGCCCTCTCTAC 1731  
QY 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
DB 1732 TTTTCTCTCTCTCTCTGTCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 1791  
QY 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
DB 1792 GATGATAGTATCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAG 1851  
QY 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
DB 1852 GTCAAGAGCCCCAGGCTTACCTTCCGAGCTACAGCGCCATCTTTACTTAATGGG 1911  
QY 621 AspPheAspGlyGlyAsnPheThrGluLeuAspAlaLysThrValThrAlaGlu 640

1912 GACTTCGATCGCGAAACTTTTATTTTCACTGAACCTGGATCCCAAGACCGTGGAGCGAGAG 1971  
641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
1972 GTGAGGCTCAGTGTGAAGAGCGGTGGAGTCTCTTCAGCACTGAAACCCACATGGA 2031  
661 ValysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680  
2032 GTGAAGCTGTTCACACGGGGGACGCGTGTGCCATGCCCTGTGTTCACCTGGACCT 2091  
681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValysMetLeuPheSerPro 700  
2092 CGACACAGCGAGCGGACAGGGTGCAGGACGATGACCTGTGAAGATGCTCTTCAGCCCA 2151  
701 GluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGluPro 720  
2152 GAAGAGATGGACCTCTCCAGGAGCAGCCCTGGATGCCAGCAGGGGCCCGGAACT 2211  
721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
2212 GCACAGAGTCTCTCAGGCACTGAATCGAAGCCCAAGGATGAGCTA 2259

RESULT 2  
ABZ11346  
ID ABZ11346 standard; cDNA; 2242 BP.  
XX  
AC ABZ11346;  
XX  
DT 20-JAN-2003 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 228.  
XX  
KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;  
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
KW antiarthritic; gene; ss.  
XX  
OS Homo sapiens.  
XX  
FN WO200270539-A2.  
XX  
PD -12-SEP-2002.  
XX  
XX 05-MAR-2002; 2002WO-US005095.  
XX  
XX 05-MAR-2001; 2001US-00799451.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
XX Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
XX Wehrman T, Wang J, Wang D, Drmanac RT;  
XX  
XX WPI: 2002-759812/82.  
XX  
XX P-PSDB; ABP69129.  
XX  
XX New polynucleotides comprising sequences assembled from expressed  
XX PT sequence tags (ESTs), useful for treating cell-proliferative,  
XX PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet  
XX PT or coagulation disorders.  
XX  
XX Claim 1; SEQ ID NO 228; 1012pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated polynucleotide (1) comprising a  
XX CC nucleotide sequence selected from any of 948 sequences (ABZ11119-  
XX CC ABZ12066) or their mature protein coding portion, active domain coding  
XX CC protein or complementary sequences. The polynucleotides are useful for  
XX CC identifying expressed genes or for physical mapping of human genome. The

CC encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight  
CC markers, as a food supplement, for generating antibodies, in medical  
CC imaging, screening and diagnostic assays and for treating cell-  
CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's  
CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,  
CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,  
CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver  
CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),  
CC arthritis, etc. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 2242 BP; 518 A; 655 C; 625 G; 444 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 0 Length: 2242  
Score: 634.00 Matches: 734  
Percent Similarity: 99.46% Conservative: 0  
Best Local Similarity: 99.46% Mismatches: 2  
Query Match: 86.14% Indels: 4  
DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x ABZ11346 (1-2242)

QY 1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaLaser 20  
DB 32 ATGCGGTACGCGCGTGAAGCTGTGACCACTGTGGTGTGCGGCCCTGCCTCC 91  
QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40  
DB 92 CAAGCGAGTGTGAGTCCGAGGAGATGGGCATGGTGACCCCTGATCTCTTCGCC 151  
QY 41 GluGlyThrAlaAlaTyrAla-ArgGlyAspTrpProGlyValValLeuSerMetGluAr 60  
DB 152 GAGGGAGCGGAGCCCTACTC-CGCGGGGACTGGCCCGGGGTGGTCTGAGCATGGAACG 210  
QY 60 gAlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgGlyCysArgThrGlnCysAl 80  
DB 211 GCGCTGTGCTCCCGGAGCGCTCCGCGCCCTTGGCTGCGCTGCGGACCACTGTCG 270  
QY 80 aAlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGl 100  
DB 271 CGCGAGCTTCCCGTGGGAGCTGAGCCCGGAGTGGTCCCCAGCCCGGCCCTCGGG 330  
QY 100 YAlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysL 120  
DB 331 CGCGCG-CGCGCTGCGCGACCTGAGCTTCTCGGGGGCTTCTGCGTGGCTGCGCTGCC 389  
QY 120 euArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluP 140  
DB 390 TGGCGGCTGCTCGGGCGCGCGCGCCACTCGCTCAGCGAAGAGATGGAGCTGGAGT 449  
QY 140 heArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuG 160  
DB 450 TCGCGAAGGGAGCCCTTACACTACTCGAGTTCGCTTCTCAAGATCAACAAGTTGG 509  
QY 160 lAlysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetG 180  
DB 510 AGAAAGCTGTGTGTCAGCAGCACACACTTCTTCTGCGGCAATCTCTGAGCATGGAATGC 569  
QY 180 lGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspL 200  
DB 570 AGCAGAACCTAGACTATTACCAACCACTGTCTGAGTGAAGAGGCGGCTTCAAGATC 629  
QY 200 euGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluG 220  
DB 630 TTGAGACTCAACCCCATATGCAAGAATTCGACTGGGAGTGGGCTCTACTCAGAGAAC 689  
QY 220 lProGlnAlaValProHisGluAlaAlaLeuGlnGluTyrPheValAlaTyrG 240  
DB 690 AGCCACAGGAGGTGTGCCCACTAGAGCGCGGCTGCAAGAAATCTTTTGGGCTATG 749  
QY 240 luGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluT 260

750 AGAGTGCCTGCTCTCGAAGGCGCCCTATGACTACGATGCCTACCACTACCTTGAGT 809  
 260 YrAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysG 280  
 810 ACACGCTGACCTCTCCAGGCCATCACAGATCATATCATCCAGTCTCTCACTGTAAGC 869  
 280 InAsnCysValThrGluLeuAlaSerHisProSerArgLysProPheGluAspPheL 300  
 870 AGAAGCTGTACAGGAGCTTGCTCCACCCCAAGTCGAGAGAAGCCCTTCAAGACTTCC 929  
 300 euProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnA 320  
 930 TCCCATCGCATTAATATATCTGCAAGTTGCTACTATAAATTTGGGAATTATACACAAG 989  
 320 laGlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
 990 CTGGTGAATGTGCCAAGACCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1049  
 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG 360  
 1050 AFTTGGCTTATTATCAGCTATGCTTGGAGAGAACACACAGATCCATCGGCCCCCGTG 1109  
 360 luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhePheAlaT 380  
 1110 AGAGTGCACAGAGTACCGACAGCGAGCCCTACTCGAAGAAAGACTGCTTTCTCTCGCTT 1169  
 380 YrAspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIleP 400  
 1170 ATGATGTTTTTGGAAATCTCTTTGTGGATTCGGATTTCATGGACTCCAGAGAAGTGATTC 1229  
 400 rGLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
 1230 CCAAGAGATTGCAGAGAAACAGAAAGTCAAGACGGGAACAGCCCTAGCATCTCCACGG 1289  
 420 luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440  
 1290 AGATTGGGAACCTTATGAAGAAATCGAGACCTTGTGGAAGAGAGAACCAAGGAGTCAC 1349  
 440 euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluLysIleSerLeuT 460  
 1350 TGGATGTGACGACAGTACCCGGGAAGTGGCCCTCTGTATGAAGGCAATCATGTCTCA 1409  
 460 hrMetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerA 480  
 1410 CCATGAATCTCCAACTCTGATGTTACAGCGGTGGTGTATGAGCGGCTATCTCTG 1469  
 480 spHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyT 500  
 1470 ACCACGAGTGTGAGAGCTGAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGCT 1529  
 500 YrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheL 520  
 1530 ACCGGGTGACAGCTCCCACTATCCATCCATGAAGATTCTATGGTGTCTACTGTCTTCA 1589  
 520 YsAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrA 540  
 1590 AAGCCCTCAAGCTGGGGCAAGAAGCAAAAGTTCTCTGCGAGAGTGCCTGTACTACTA 1649  
 540 snValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuT 560  
 1650 ACGTACGAGAGAGAGTGGCGGCATCATGGAGTCTTACTTCCGCTGGATACGCCCTCT 1709  
 560 YrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgL 580  
 1710 ACTTTTCTACTCTCATCTGCTGCGCAGTCCATCGAAGAGGTCCAGGACAGAGGA 1769  
 580 YsAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValC 600  
 1770 AGGATGATAGTATCCAGTCCAGTGGACAACTGCATCTCTGAATGCCGACACCTCTGT 1829  
 600 YsValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnG 620

1830 GTGTCAAAGAGCCCCCAGCCTACACCTTCGCGACTACAGCGCCATCTTTTACCTAAATG 1889  
 620 lyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaG 640  
 1890 GGGACTTCGATGCGGAAACTTTTATTTCTACTGAATGCAAGACCGTGACGGCAG 1949  
 640 luValGlnProGlnCysGlyArgAlaValIlePheSerSerGlyThrGluAsnProHisG 660  
 1950 AGTGCAGCCTCAGTGTGGAAGAGCCGTGGATTCTTTTCAGGCACTGAAACCCACATG 2009  
 660 lyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAsp 680  
 2010 GAGTGAAGGCTGTACACAGGGGACGCTGTGCCATCGCCCTGTGTTCACCCCTGGACC 2069  
 680 roArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerP 700  
 2070 CTCGACACAGGAGCGGACAGGGTGCAGGACAGTACCTGGTGAAGATGCTCTTCAGCC 2129  
 700 roGluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGluP 720  
 2130 CAGAAGAGATGGACCTCTCCAGGACAGCCCTGGATCCCGAGGGGCCCCCGGAC 2189  
 720 roAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
 2190 CTGCACAGAGTCTCTCTCAGGACAGTGAATCGAAGCCCAAGGATGAGCTA 2239  
 RESULT 3  
 AAS17573  
 ID AAS17573 standard; cDNA; 2211 BP.  
 XX  
 AC AAS17573;  
 DX 26-FEB-2002 (first entry)  
 DE DNA encoding novel secreted protein #2.  
 XX  
 KW Secreted protein; cytostatic; immunosuppressive; vulnery; vaccine;  
 anti-inflammatory; neuroprotective; nephrotropic; cardiovascular; human;  
 cancer; autoimmune disease; wound healing disorder; infection;  
 haematopoietic disorder; inflammatory disorder; infertility;  
 neurological disease; psychiatric disease; cardiovascular disease;  
 respiratory disease; renal; gastrointestinal; ss.  
 XX  
 OS Homo sapiens.  
 Key Location/Qualifiers  
 CDS 1..2211  
 /tag= a  
 /product= "Human secreted protein"  
 XX  
 PN MO200179454-A1.  
 XX  
 PD 25-OCT-2001.  
 XX  
 PF 11-APR-2001; 2001WO-US011797.  
 XX  
 PR 13-APR-2000; 2000US-0196603P.  
 PR 24-APR-2000; 2000US-0199417P.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 PI Agarwal P, Murdock PR, Rizvi SK, Smith RF, Xiang Z;  
 DR WPI; 2002-061975/08.  
 DR P-PSDB; AAU09861.  
 XX  
 PT New secreted proteins or polypeptides, useful for treating e.g. cancer,  
 autoimmune diseases, wound healing disorder, infections, hematopoietic  
 disorders, inflammatory disorders, infertility, cancer.  
 XX  
 PS Claim 2; Page 35-36; 92pp; English.

CC The invention relates to an isolated novel secreted polypeptide (I) and  
 CC polynucleotide (II). (I) and (II) are useful for treating cancer,  
 CC autoimmune diseases, wound healing disorder, infections, haematopoietic  
 CC disorders, inflammatory disorders, infertility, neurological and  
 CC psychiatric diseases, cardiovascular diseases, respiratory diseases,  
 CC renal diseases, or gastrointestinal diseases. These may also be used to  
 CC treat diseases, abnormalities and disorders caused by abnormal  
 CC expression, production, function and/or metabolism of the genes, as  
 CC vaccines for inducing immunological response in a mammal, and in  
 CC screening methods for detecting the effect of added compounds on the  
 CC production of mRNA and polypeptide in cells. The polypeptides can be used  
 CC as immunogens to produce antibodies immunospecific for the polypeptides,  
 CC and to identify membrane-bound or soluble receptors. The polynucleotides  
 CC may be used as diagnostic reagents, in chromosome localisation studies,  
 CC and in tissue expression studies. The present sequence represents the  
 CC coding sequence of novel human secreted protein #2

XX Sequence 2211 BP; 512 A; 549 C; 612 G; 438 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.: 0 Length: 2211  
 Score: 433.00 Matches: 733  
 Percent Similarity: 99.19% Conservative: 0  
 Best Local Similarity: 99.19% Mismatches: 3  
 Query Match: 58.83% Indels: 6  
 DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x AAS17573 (1-2211)

QY 1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaAsp 20  
 Db 1 ATGGCGGTACGGGGTGAAGCTGCTGACCATCTGCTGGCTGCTGGCGCTGCCCTCC 60  
 QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40  
 Db 61 CAAGCCGAGGTCGAGTCCGAGCAGGATGGGCATGGTACGCTGTCTCTTCGCC 120  
 QY 41 GluGlyThrAlaAlaTyraAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
 Db 121 GAGGGACCGCAGCCCTACGCGCGGGGACCTGGCCGGGGTGGCTTGGAGCATGGAACGG 180  
 QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysAysThrGlnCysAla 80  
 Db 181 GCGTCTCGCTCCCGGAGCCCTCCGCGCCCTTCGCTCGCTCGCTCGCCACCCAGTGTGCC 240  
 QY 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
 Db 241 GCGGACTTCCCGTGGAGCTGGACCCCGACCTGGTCCCGCCAGCCCGGCCCTCGGGC 300  
 QY 101 AlaGly-AlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgArgAlaAlaCysLe 120  
 Db 301 GCGGC-CGCCCCCGCGACCTGAGCTTTTCGGGGGCGCTTTCGGCTCGCTCGCTGCCCT 359  
 QY 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140  
 Db 360 GCGCGCTGCTCGGGCGCGCGCGCCGACCTCGCTCAGCGAAGAGATGGAGCTGGAGTT 419  
 QY 140 eArgLysArgSerProTyraAsnTyraLeuGlnValAlaTyraPheLysIleAsnLysLeuG1 160  
 Db 420 CCGCAAGCGGAGCCCTCAACTACTCTGAGTTCGCTTCAAGATCAACAGTTGGA 479  
 QY 160 uLysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetG1 180  
 Db 480 GAAGCTGTGTGTCAGACACACACTTCTTCGTGGGCAATCTGAGCACATGGAATGCA 539  
 QY 180 nGlnAsnLeuAspTyraTyraThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
 Db 540 GCAGAACCTAGACTATTACCAACCATGCTGGAGTGAAGGAGGCCGACTTCAAGGATCT 599  
 QY 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyraSerGluGluG1 220  
 Db 600 TGAGACTCAACCCCATATCAAGAATTTCGACTGGGAGTGGCAGCTCTACTCAGAGAAACA 659

QY 220 nProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTyraPheValAlaTyraG1 240  
 Db 660 GCCACAGGAAGCTGTGCCCCACCTAGAGCGCGCTGCAAGAATACTTTGTGGCCCTAGA 719  
 QY 240 uGluCysArgAlaLeuCysGluGlyProTyraAspTyraAspGlyTyraAsnTyraLeuGluTy 260  
 Db 720 GGAGTGCCGTGCCCTCTCGGAAGGCGCTTATGACTACCATGGCTACCACTACCTTGAGTA 779  
 QY 260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyraIleGlnValLeuAsnCysGlyG1 280  
 Db 780 CAACGCTGACCTCTTCCAGGCCATCACAGATCATTTACATCCAGGTCCTCACTGAAGCA 839  
 QY 280 nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
 Db 840 GAACCTGTGTACGGAGCTTGTCTCCACCCCAAGTCGAGAGAAGCCCTTTTGAAGACTTCT 899  
 QY 300 uProSerHisTyraAsnTyraLeuGlnPheAlaTyraTyraAsnIleGlyAsnTyraThrGlnAl 320  
 Db 900 CCCATCGCATTAATATCTCGAGTTTGCCTACTATTAACATTGGGAATTAACACAGGC 959  
 QY 320 aGly-GluCysAlaLysThrTyraLeuPhePheProAsnAspGluValMetAsnGlnA 340  
 Db 960 TGT-TGAATGTGCCAAGACCTATCTCTCTTCTCCCAATCAGCAGGTGATGAACCAAA 1018  
 QY 340 snLeuAlaTyraTyraAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG 360  
 Db 1019 ATTTGCGCTATTATGCGATGCTTGTGGAGAAGAACACACACAGATCCATCGGCCCGCTG 1078  
 QY 360 lSerAlaLysGluTyraArgGlnArgSerLeuLeuGlyLysGluLeuLeuPhePheAlaT 380  
 Db 1079 AGAGTCCCAAGAGTACCGACAGCGAAGCTTACTTGGAAAAAGAACTGCTTTTCTTCGCTT 1138  
 QY 380 rAspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIleP 400  
 Db 1139 ATGATGTTTTGGAAATCCCTTTGTGATCCGGATTCATGGACTCCAGAAAGAGTATTC 1198  
 QY 400 rLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
 Db 1199 CCAAGAGATTGCAAGAGAAACAGAGTCAAGCGGAAACAGCGGTACGCATCTCCAGG 1258  
 QY 420 lIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440  
 Db 1259 AGATTCCGAACCTTATGAAGGAAATCGAGACCTTGTGGAAGAGAAGACCAAGAGTCA 1318  
 QY 440 euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyraGluGlyIleSerLeuT 460  
 Db 1319 TGATGTGACAGACTGACCCGGAAGTGGCCCTCTGCTGTATGAAGCATCAGTCTCA 1378  
 QY 460 hrMetAsnSerLysLeuLeuAsnGlyTyra-GlnArgValValMetAspGlyValIleSer 479  
 Db 1379 CCATGAACCTCAAACTCCTGAATGGTTC-CCAGCGGGTGGTGTATGGACGCGGTAATCTCT 1437  
 QY 480 AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499  
 Db 1438 GACCACAGTGTGAGAGCTGACAGACTGACCAATGTGGCAGCAACCTCAGAGATGGC 1497  
 QY 500 TyraGlyGlnThrSerProHisThrProAsnGluLysPheTyraGlyValThrValPhe 519  
 Db 1498 TACCGGGGTGACAGCTCCCGACATCTCCCAATGAAAGTTCTATGGTGTGCTGCTCTTC 1557  
 QY 520 LysAlaLeuLysLeuGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyraTy 539  
 Db 1558 AAAGCCCTCAAGCTGGGCGCAAGGCAAGATTCCTCTGCAAGTGGCCCACTGTACTAC 1617  
 QY 540 AsnValThrGluLysValArgArgIleMetGluSerTyraPheArgLeuAspThrProLeu 559  
 Db 1618 AACGTGACGAGAGAGTGGCGGCATCATGAGTCTTACTTCCGCTGGATAGCGCCCTC 1677  
 QY 560 TyraPheSerTyraSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArg 579  
 Db 1678 TACTTTTCTTACTCTCATCTGGTGGCGCACCTGCCATCGAAGAGGTCCAGGAGAGG 1737  
 QY 580 LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599





Db 630 TGAGACTCAACCCCATATGCAAGATTTCGACTGGGAGTCGGACTCTACTACAGGAACA 589  
Qy 220 nProGlnGluAlaValProHisLeuGluAlaLeuLeuGlnGluTyrPheValAlaTyrGl 240  
Db 690 GCCACAGAACTGTGCCCCACCTAGAGCGCGCTGCAAGATACTTTGTGGCCCTATGA 749  
Qy 240 uGluCysArgAlaLeuCysGluGluProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
Db 750 GGAGTGGCCGTGCCCTTCGCAAGGCCCTATGACTAGATGGCTCAACACTACCTTACCTG 809  
Qy 260 rAsnAlaAspLeuPheGlnAlaLeuThrAspHisTyrIleGlnValLeuAsnCysLysGl 280  
Db 810 CAACGCTGACCTCTTCAGGCCATCACAGATCAATACATCCAGTCTCAACTGTAAACA 869  
Qy 280 nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
Db 870 GAACGTGTCCAGGAGCTTGTCTCCACCCCAAGTCGAGAGGCCCTTTCAAGACTTCCT 929  
Qy 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320  
Db 930 CCCATCCGATATATATATCTGCAAGTTTGGCTACTATAACATTGGGAATTATACAGGC 989  
Qy 320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
Db 990 TGT-TGAATGTGCCAAGACCTATCTTCTCTTCCCCCAATGACGAGGTGATGAACCAA 1048  
Qy 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG 360  
Db 1049 ATTTGGCCCTATTATGCACTATGCTTGGAGAAGAACACACAGATCCATCGGCCCGCTG 1108  
Qy 360 luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAla 380  
Db 1109 AGAGTGCACAGAGTACGACAGCAAGCCCTACTGGAAAGAACTCTCTTTCTCGCTT 1168  
Qy 380 yrAspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIleP 400  
Db 1169 ATGATGTTTTTGGAAATCCCTTTCTGGATCCGGATTTCATGCACTCCAGAGAAAGTGATTC 1228  
Qy 400 roLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
Db 1229 CCAAGAGATTCCAGAGAAACAGAAAGTCAGAACCGGAAACAGCCGTACGATCTCCCAAG 1288  
Qy 420 luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440  
Db 1289 AGATTGGGAACCTTATGAAGGAATCGAGACCCCTTGTGGAAGAGAACCAAGAGGTAC 1348  
Qy 440 euAspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyIleSerLeuT 460  
Db 1349 TCGATGTGAGCAGACTGACCCGGAAGGTGCCCCCTCTGTATGAAGGCATCAGTCTCA 1408  
Qy 460 hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer 479  
Db 1409 CCATGAATCCAACTCTCTGATGTTCC-CCAGCGGGTGGTGTGATGGAGCGGTAATCTCT 1467  
Qy 480 AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499  
Db 1468 GACCACGAGTGTACGAGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGC 1527  
Qy 500 TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe 519  
Db 1528 TACCGGGGTACAGCTCCCACTACTCCCAATGAAAGTTCTATGTGTCTCTCTCTCTC 1587  
Qy 520 LysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyr 539  
Db 1588 AAAGCCCTCAAGCTGGGGCAAGAGGCAAGTTCTCTGCAAGGTGCCCACTGTACTAC 1647  
Qy 540 AsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeu 559  
Db 1648 AACGTGACGGAAGGTGGCGCATCATGAGTCTTCTTCCGCTCGATACGCCCTC 1707  
Qy 560 TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArg 579  
Db 1708 TACTTTTCTACTCTCTGTTGTCGCGCACTGCATCGAAGAGTCCAGCGCAGAGAGG 1767

Qy 580 LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599  
Db 1768 AAGGATGATAGTCATCCAGTCCAGTGGACAACTGCATCTCTGAATGCCGAGACCTCTGTG 1827  
Qy 600 CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619  
Db 1828 TGTGTCAAAGAGCCCGGAGCTTACCTTCCGGACTACAGGCCATCTTTACCTAAAT 1887  
Qy 620 GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla 639  
Db 1888 GGGGACTTCGATGGCGGAAACTTTTATTTTCACTGAACCTGGATGCCAAGACCGTGACGCA 1947  
Qy 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659  
Db 1948 GAGGTGACCTTCAGTGTGAAGAGCCGTGGGATTTCTTTCAGGCATCGAAACCCACAT 2007  
Qy 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAsp 679  
Db 2008 GGAGTGAAGCTGTCCACAGGGGCGAGCGCTGTGCATCGCCCTGTGGTTCACCTGGAC 2067  
Qy 680 ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer 699  
Db 2068 CCTCGACACAGCGAGCGGACAGGGTGCAGGCAGATGACCTGGTGAAGATGCTCTTCAGC 2127  
Qy 700 ProGluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGlu 719  
Db 2128 CCAGAAAGATGGACCTCTCCAGGAGCAGCCCTGGATGCCAGAGGGCCCCCGGAA 2187  
Qy 720 ProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2188 CTTGCACAAAGAGTCTCTCTCAGGCAGTGAATCGAAGCCCAAGGATGAGCTA 2238

## RESULT 5

AAH14589 standard; cDNA; 2585 BP.

AAH14589;

26-JUN-2001 (first entry)

Human cDNA sequence SEQ ID NO:12193.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000BP-00116126.

29-JUL-1999; 99JP-00248036.

27-AUG-1999; 99JP-00300253.

11-JAN-2000; 2000JP-00118776.

02-MAY-2000; 2000JP-00183767.

09-JUN-2000; 2000JP-00241899.

(HELI-) HELIX RES INST.

WPI; 2001-318749/34.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-

length cDNAs defined in the specification, and for the detection and/or

diagnosis of the abnormality of the proteins encoded by the full-length

cDNAs.

Claim 8; SEQ ID NO 12193; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to a complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AA892446 to AA895893 represent human amino acid sequences; and AAH13629 to AAH13632 represent polynucleotides, all of which are used in the exemplification of the present invention.

Sequence 2585 BP; 590 A; 759 C; 723 G; 513 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0	Length:	2585
Score:	433.00	Matches:	733
Percent Similarity:	99.19%	Conservative:	0
Best Local Similarity:	99.19%	Mismatches:	3
Query Match:	58.83%	Indels:	6
DB:	4	Gaps:	0

US-10-045-815-4 (1-736) X AAH14589 (1-2585)

QY	1	MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValValAlaAlaSer	20
DB	53	ATGGGGGTATCGCGCTTGAAGCTGTCACACACTGTGCTGCTCGTGGCGCTGCCTCC	112
QY	21	GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla	40
DB	113	CAAGCGGAGGTCCAGTCCGAGCGAGAGTGGGCGATGGTGACGCTGATCTGCTCTTCGCC	172
QY	41	GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg	60
DB	173	GAGGGGACCGCAGCCTACGCGCGGGGACTGGCGCCGGGGTGCTCTGAGCATGGAAACGG	232
QY	61	AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgGluLeuArgCysArgThrGlnCysAla	80
DB	233	CGCGTGCCTCCCGGGCAGCCCTCCGCGCCCTTCGCGTGGCGCACCCAGATGTGCC	292
QY	81	AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly	100
DB	293	GCCGACTTCCCGTGGAGCTGGACCCGACTGGTGCCCCAGCCGCCGAGCCTCGGGC	352
QY	101	AlaGly-AlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgArgAlaAlaCysLe	120
DB	353	CGCGC- CGCCCTCGCGACCTCAGCTTCCTTCGGGGGCTTCTCGTTCGCGCTGCCTGCCT	411
QY	120	uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluLeuMetGluLeuGluPhe	140
DB	412	CGCCGCTTGCCTCGGGCGCGCGCGCCCACTCGCTCAGCGAGAGATGGAGCTGGAGTT	471
QY	140	eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGln	160
DB	472	CGCAAGCGGAGCGCCCTACAACTACCTGCAGGTTCGCTACTTCAAGATCAACAAGTTGGA	531
QY	160	uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln	180
DB	532	GAAGCTCTTCTCGAGCACACCTTCTTCGTGGGCAATCTCTGAGCATCGAATAATGCA	591
QY	180	nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe	200

```
QY 560 TyrPheSerTyrSerHisLeuValCysArgThrAlaIleLeuGluValGlnAlaGluArg 579
Db 1730 TACTTTCTCTACTCTCATCTGGTGGCGCACTGCCATCGAAGAGGTCCAGCGAGAGG 1789
QY 580 LysAspSerHisProValHisValAspAsnCysIleLeuAenAlaGluThrLeuVal 599
Db 1790 AAGGATGATGATCATCGATCCAGTGCAGCAACTGCATCTCTGAATGCCGAGACCCCTCGTG 1849
QY 600 CysValIysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619
Db 1850 TGTGTCAAAGAGCCCCAGACGCTACACCTTCGCGACTACAGCGCCATCCITTACTTAAT 1909
QY 620 GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaIysThrValThrAla 639
Db 1910 GGGGACTTCGATGGCGGAACCTTTTATTTCACTCAACTGGATGCCAAGCCGTGACGGCA 1969
QY 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659
Db 1970 GAGGTGAGCTCTAGTGTGGAGAGCGGTGGATTCCTTCAGGCATCTGAACCCACAT 2029
QY 660 GlyValIysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAsp 679
Db 2030 GGAGTGAAGGCTGTCAACAGGGGCGAGCGCTGTGCCATCGCCCTGTGGTTCAACCTGGAC 2089
QY 680 ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValIysMetLeuPheSer 699
Db 2090 CCTCGACACAGCGAGCGGAGCAGGCTGCAGCAGATGACCTGTGGAGATGCTCTTCAGC 2149
QY 700 ProGluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGlu 719
Db 2150 CCAGAAGAGATGGACCTCTCCAGAGAGCAGCGCCCTGGATGCCAGCAGGGCCCCCGAA 2209
QY 720 ProAlaGlnGluSerLeuSerGlySerGluSerIysProIysAspGluLeu 736
Db 2210 CCTGCACAAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2260
RESULT 6
AAH14513
ID AAH14513 standard; cDNA; 2524 BP.
XX
AC AAH14513;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:12044.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EPI074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length
```

```
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 12044; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dr primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
```

Sequence 2524 BP; 583 A; 747 C; 704 G; 490 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0	Length:	2524
Score:	407.00	Matches:	707
Percent Similarity:	99.16%	Conservative:	0
Best Local Similarity:	99.16%	Mismatches:	3
Query Match:	55.30%	Indels:	6
DB:	4	Gaps:	0

US-10-045-815-4 (1-736) x AAH14513 (1-2524)

```
QY 1 MetAlaValArgAlaLeuLysLeuThrThrLeuAlaValValAlaAlaSer 20
Db 16 ATGGCGGTACGGCGGTGAAGCTGCTGACACACACTGCTGGCTGTGGTCCGCTCC 75
QY 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuPheAla 40
Db 76 CAAGCGAGTTCAGTCCGAGCAGAGTGGGCGATGGTGGCTGATCTGCTTCGCC 135
QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60
Db 136 GAGGGGACCGGACCTACGCGCGCGGGAGTGGCGCGGCTGGCTGGCTGGCTGGCTGG 195
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
Db 196 GCGCTGCGCTCCCGGCGAGCCCTCCGCGCCCTTCGCTGGCTGGCTGGCTGGCTGGCT 255
QY 81 AlaAspPheProTyrGluLeuAspProAspTyrProSerProAlaGlnAlaSerGly 100
Db 256 GCGGACTTCCGCTGGAGCTGGACCCGACTGGTCCCGCAGCGCGCCGAGCCCTCGGC 315
QY 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyClyLeuLeuArgAlaAlaCysLe 120
Db 316 GCGCGC-CGCGCTGCGCGACCTGAGCTTCTTCGGGGCGCTTCGCTGGCTGGCTGGCT 374
QY 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140
Db 375 GCGCGCTGCTTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 434
QY 140 eArgIysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheIysIleAsnLysLeuGl 160
Db 435 CGGCAAGCGGAGCGCCCTCACTACTACCTGAGTCCGCTGCTGCTGCTGCTGCTGCTGCT 494
QY 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGl 180
Db 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGl 180
```

495 GAAAGCTGTCGTCAGCAGACACACCTTCTCGTGGCAATCCTCGAGCAGCATGGAAATGCA 554  
180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
555 GAGAACTCTAGACTATTACCAACCATGCTCTGGAGTGAAGAGCCGACATTCAGAGATCT 614  
200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluG 220  
615 TCAGACTCAACCCCATATGCAAGAAATTCGAGCTGGAGTGGAGCTACTACTCAGAGGAACA 674  
220 nProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTyrPheValAlaTyrG 240  
675 GCCACGAGAGCTGTGCCCCACCTAGAGCGCGCTGCAAGAACTCTTGTGGCTATGA 734  
240 uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluT 260  
735 GGAGTGCCGTGCTCTCGAAGGCCCTATGACTACGATGGCTACAACTACCTTTGAGTA 794  
260 rAsnAlaAspLeuPheGlnAlaLeuThrAspHisTyrIleGlnValLeuAsnCysLysG 280  
795 CAACGCTGACCTCTTCCAGGCGCATCACAGATCATTACATCCAGGTCTCAACTGTAAACA 854  
280 nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
855 GAACCTGTCTACGGAGCTTGCTTCCACCAAGTCGAGAGAGCCCTTTGAAGACTTCCT 914  
300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320  
915 CCATCGCATTAATAATTATCTGCAGTTTGCCTACTATAAATGGGAATTATACACAGGC 974  
320 sGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
975 TGT-TGAATGTGCCAAGACCTATCTTCTCTTCCCAATGACGAGGTGATGAAACCAA 1033  
340 sNLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArg 360  
1034 ATTTGGCTATTATGACGCTATGCTTTGGAGAGAAACACACAGATCCATCGCCCCCGTG 1093  
360 uSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAla 380  
1094 AGAGTGCCCAAGAGTACCGACAGCGAGCCCTACTGGAAGAAAGAACTGCTTTCTTCGCTT 1153  
380 yAspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIleP 400  
1154 ATGATGTTTGGAAATTCCTTTTGGATCCGGATTCATGACTCCAGAGAGATGATTC 1213  
400 rLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
1214 CCAAGAGATTGCAAGAGAAACAGAAAGTCAAGACCGGAAACAGCGTACGATCTCCGAGG 1273  
420 lLleGlyAsnLeuMetLysGluLleGluThrLeuValGluGluLysThrLysGluSerL 440  
1274 AGATTGGGAACCTTATGAAGGAATCGAGACCTTGTGGAAGAGAGACCAAGGAGTAC 1333  
440 eAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuT 460  
1334 TGGATGTGAGCAGACTGACCCGGGAAGGTGGCCCCCTCTGTATGAAGGCATCATGTCTCA 1393  
460 hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer 479  
1394 CCATGAATCTCCAACTCTGTAATGTTTC-CCAGCGGGTGGTGTGAGCGGGGTATCTCT 1452  
480 AspHisGluCysGlnGluGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499  
1453 GACCAAGAGTGTGAGGAGCTGACAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGC 1512  
500 TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe 519  
1513 TACCGGGGTGACAGCTCCCCACATACTCCCAATGAAAGTTCTATGGTGTCTACTGTCTTC 1572  
520 LysAlaLeuLysLeuGlyGlnGlyLysValProLeuGlnSerAlaHisLeuTyrTyr 539  
1573 AAAGCCCTCAAGCTGGGGCAAGAGCAAGATTCCTCTGAGAGTGGCCCACTGTACTAC 1632

QY 540 AsnValThrGluLysValArgIleMetGluSerTyrPheArgLeuAspThrProLeu 559  
Db 1633 AACGTGACGAGAGGTGGCGGCATCATGGAGTCTACTTCCGCTGGATACGCCCTC 1692  
QY 560 TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGluAlaGluArg 579  
Db 1693 TACTTTTCTACTCTCACTGTGTGGCGCATGTCATCGAAGAGTCCAGCAGAGAGG 1752  
QY 580 LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599  
Db 1753 AAGGATGATAGTATCCAGTCCAGCTGGACAACTCATCTGAATGCCAGAGACCTCGTG 1812  
QY 600 CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619  
Db 1813 TGTGTCAAAGAGCCCCAGCCTACACCTTCCCGACTACAGGCCATCCTTTACCTAAT 1872  
QY 620 GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla 639  
Db 1873 GGGGACTTCGATGGCGGAACCTTTATTCTCACTGAATGGATGCCAGACCGTGACGCA 1932  
QY 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659  
Db 1933 GAGTGCACCTCAGTGTGGAAGAGCCGTGGGATTCCTCTCAGGCACATGAAACCCACAT 1992  
QY 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAsp 679  
Db 1993 GGAGTGAAGCTGTACCGAGGGGCGAGCTGTGCTGCTCGCCCTGTGTTTCACTCCCTGGAC 2052  
QY 680 ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer 699  
Db 2053 CCTCGACACAGCGAGCGGACAGGCTGCAGGCAGATGACCTGCTGAAGATGCTCTTTCAGC 2112  
QY 700 ProGluGluMetAspLeuSerGlnGluGlnPro 710  
Db 2113 CCAGAGAGATGACCTCTCCAGAGAGGAGCC 2145  
RESULT 7  
AAD08489  
ID AAD08489 standard; cDNA; 2577 BP.  
XX AAD08489;  
AC AAD08489;  
DT 09-AUG-2001 (first entry)  
XX Human secreted protein-encoding gene 2 cDNA clone HUVFY29, SEQ ID NO:12.  
DE Human; secreted protein; proliferative disorder; cancer; tumour;  
XX foetal abnormality; developmental abnormality; haematopoietic disorder;  
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;  
KW endocrine disorder; infection; wound healing; vulnery; cell culture;  
KW chemotaxis; food additive; binding partner identification; ss.  
XX Homo sapiens.  
OS  
XX Key Location/Qualifiers  
XX CDS 24..1103  
FT /tag= a  
FT /product= "Human secreted protein precursor"  
FT sig\_peptide 24..77  
FT /tag= b  
FT mat\_peptide 78..1100  
FT /tag= c  
FT /product= "Mature human secreted protein"  
XX W0200136432-A2.  
XX 25-MAY-2001.  
PD

XX		15-NOV-2000; 2000WO-US031162.
PF		
XX		
XX		19-NOV-1999; 99US-0166415P.
PR		30-JUN-2000; 2000US-0215136P.
XX		
XX		(HUMA-) HUMAN GENOME SCI INC.
PA		
XX		
PI	Ruben SM, Komatsoulis GA, Baker KP, Young PE;	
XX		
DR	WPI; 2001-343793/36.	
DR	P-PSDB; AAE04200.	
XX		
XX		
PT	Isolated nucleic acid molecule encoding a human secreted protein is used	
PT	in preventing, treating or ameliorating a medical condition.	
XX		
PS	Claim 1; Page 403-404; 509pp; English.	
XX		
CC	AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted	
CC	protein genes, and AAE04199-AAE04239 represent the proteins they encode.	
CC	AAE04240-AAE04297 represent human secreted protein fragments or variants.	
CC	The secreted proteins and their genes are useful for preventing, treating,	
CC	or ameliorating medical conditions, e.g., by protein or gene therapy.	
CC	Pathological conditions can be diagnosed by determining the amount of the	
CC	new protein in a sample or by determining the presence of mutations in	
CC	the new genes. Specific uses are described for each of the 18 genes.	
CC	based on the tissues in which they are most highly expressed, and include	
CC	developing products for the diagnosis or treatment of proliferative	
CC	disorders, cancer, tumours, foetal and developmental abnormalities,	
CC	haematopoietic disorders, diseases of the immune system, AIDS, autoimmune	
CC	diseases (e.g., rheumatoid arthritis), inflammation, allergies,	
CC	neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),	
CC	cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,	
CC	psoriasis) sepsis, diabetes, atherosclerosis, cardiovascular disorders,	
CC	angiogenic disorders, kidney disorders, gastrointestinal disorders,	
CC	pregnancy-related disorders, endocrine disorders, and infections. The	
CC	proteins can also be used to aid wound healing and epithelial cell	
CC	proliferation, to prevent skin ageing due to sunburn, to maintain organs	
CC	before transplantation, for supporting cell culture of primary tissues,	
CC	to regenerate tissues, to identify their cognate ligands or binding	
CC	partners, and in chemotaxis, and can be used as a food additive or	
CC	preservative to modify storage properties. Antibodies specific for a	
CC	protein of the invention can be used in alleviating symptoms associated	
CC	with the disorders mentioned above, and in diagnostic immunoassays e.g.,	
CC	radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The	
CC	present sequence represents a human secreted protein-encoding cDNA of the	
CC	invention	
XX		
SQ	Sequence 2577 BP; 610 A; 753 C; 711 G; 503 T; 0 U; 0 Other;	
	Alignment Scores:	
	Pred. No.: 0 Length: 2577	
	Score: 407.00 Matches: 733	
	Percent Similarity: 98.92% Conservative: 0	
	Best Local Similarity: 98.92% Mismatches: 3	
	Query Match: 55.30% Indels: 8	
	DB: 4 Gaps: 0	
	US-10-045-815-4 (1-736) x AAD08489 (1-2577)	
Qy	1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaSer 20	
Dd	24 ATGGCGGTACGCGCTTGAGCTGCTGACCACACTGCTGGCTGTCGGCGGTGCCTCC 83	
Qy	21 GlnAlaGluValGlnSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40	
Dd	84 CAAGCCAGTCCAGTCCGAGCAGATGGGGCATGGTGACGCCCTGATCTGCTCTTCGCC 143	
Qy	41 GluGlyThrAlaAlaTyraAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60	
Dd	144 GAGGGACCGCAGCTACCGCGCGGGGACTGGCCCGGGTGGTCTTGACATCGGAACGG 203	
Qy	61 AlaLeuArgSerArgAlaAlaLeuAlaLeuArgAlaLeuArgGlySerCysArgThrGlnCysAla 80	

Db 1282 GGAGATTGGGAACCTTATGAGGAAATCGAGACCCCTTGTGAAGAGAACACCAAGGAGTC 1341  
Qy 439 rleuaspvalserargleuthrarggluglyglyproleuthrarggluglylleSerLe 459  
Db 1342 ACTGGATGTGACGACACTGACCCGGGAAGGGTGGCCCCCTGCTGATGAGGCATCAGTCT 1401  
Qy 459 uThrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIles 479  
Db 1402 CACCATGACTCCAACTCCTGAATGGTTC-CCAGCGGGTGGTATGACGCGGTAATCT 1460  
Qy 479 erAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspG 499  
Db 1461 CTGACCACGAGTGTGAGAGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGAGATG 1520  
Qy 499 lYtYrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValP 519  
Db 1521 GCTACCGGGTGCAGACTCCCACTACATCTCCATGAAAGTTCTATGGTGTCACTGCT 1580  
Qy 519 heLysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrT 539  
Db 1581 TCAAAGCCCTCAAGTGGGCAAGAGGCAAGGTTCTCTGCAGAGTGCACCACTGTACT 1640  
Qy 539 yrAsnValThrGluLysValArgArgGileMetGluSerTyrPheArgLeuAspThrProL 559  
Db 1641 ACACGTGACGAGGAGGTGGGCGCATCATGGAGTCTCTTCGGCTGGATACGCCCC 1700  
Qy 559 euTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluA 579  
Db 1701 TCTACTTTTCTACTCTCATCTGTGTGTCGCGCACTGCATCGAAGAGTCCAGGCAGAGA 1760  
Qy 579 rgLysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuV 599  
Db 1761 GGAAGGATGATAGTCATCCAGTCCACGTGGACACTGCATCTCTGAATGCCGAGACCCCTCG 1820  
Qy 599 aiCysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuA 619  
Db 1821 TGTGTGTCAAGAGACCCCGCCACTACCTTCGCGCACTACAGCGCCATCTTTTACCTAA 1880  
Qy 619 enGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrA 639  
Db 1881 ATGGGGACTTCGATGGCGGAAACTTTTATTTTCACTGAACCTGCATGCCAGACCCGTGACGG 1940  
Qy 639 laGluValGlnProGlnCysGlyArgAlaValAlaGlyPheSerSerGlyThrGluAsnProH 659  
Db 1941 CAGAGTGCAGCCTCAGTGTGGAGAGCGGTGGGATTTCTTTCAGGCACCTGAAACCCAC 2000  
Qy 659 isGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuA 679  
Db 2001 ATGGAGTGAAGCTGTCCACAGGGGGCGACGCTGTGCATCGCCTGTGTGTTTCACTCCCTGG 2060  
Qy 679 spProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheS 699  
Db 2061 ACCCTCGACACAGCAGCGCGGACAGGGTGCAGAGATGACTGCTGGTGAAGATGCTCTTCA 2120  
Qy 699 erProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProG 719  
Db 2121 GCCCAGAAGATGACCTCTCCAGGAGCAGCCCTGGATGCCAGCAGCGGCCGCCGCCG 2180  
Qy 719 luProAlaGlnLysLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2181 AACCTGCAAGAGTCTCTCAGGACGTGAATCGAAGCCCAAGATGAGCTA 2233

## RESULT 8

AAD08507

ID AAD08507 standard; cDNA; 2615 BP.

AC AAD08507;

XX AAD08507;

XX 09-AUG-2001 (first entry)

DE Human secreted protein-encoding gene 2 cDNA clone HUVFY29, SEQ ID NO:30.

XX Human; secreted protein; proliferative disorder; cancer; tumour;

XX Human; secreted protein; proliferative disorder; cancer; tumour;

foetal abnormality; developmental abnormality; haematopoietic disorder;  
immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
inflammation; allergy; neurological disorder; Alzheimer's disease;  
Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
cardiovascular disorder; angioecnic disorder; kidney disorder;  
gastrointestinal disorder; pregnancy-related disorder; gene therapy;  
endocrine disorder; infection; wound healing; vulnerability; cell culture;  
chemotaxis; food additive; binding partner identification; ss.  
Homo sapiens.  
Key Location/Qualifiers  
CDS 24..545  
FT /tag= a  
FT /product= "Human secreted protein precursor"  
FT sig\_peptide 24..77  
FT /tag= b  
FT mat\_peptide 78..542  
FT /tag= c  
FT /product= "Mature human secreted protein"  
XX WO200136432-A2.  
XX 25-MAY-2001.  
XX 15-NOV-2000; 2000WO-US031162.  
XX 19-NOV-1999; 99US-0166415P.  
XX 30-JUN-2000; 2000US-0215136P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Ruben SM, Komatsoulis GA, Baker KP, Young PE;  
XX WPI; 2001-343793/36.  
XX P-PSDB; AAE04217.  
XX Isolated nucleic acid molecule encoding a human secreted protein is used  
XX in preventing, treating or ameliorating a medical condition.  
XX Claim 1; Page 416-417; 509pp; English.

AAD08488-AAD08529 represent cDNAs corresponding to 18 human secreted  
protein genes, and AAE04199-AAE04239 represent the proteins they encode.  
AAE04240-AAE04297 represent human secreted protein fragments or variants.  
The secreted proteins and their genes are useful for preventing, treating  
or ameliorating medical conditions, e.g., by protein or gene therapy.  
Pathological conditions can be diagnosed by determining the amount of the  
new protein in a sample or by determining the presence of mutations in  
the new genes. Specific uses are described for each of the 18 genes,  
based on the tissues in which they are most highly expressed, and include  
developing products for the diagnosis or treatment of proliferative  
disorders, cancer, tumours, foetal and developmental abnormalities  
haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
angiogenic disorders, kidney disorders, gastrointestinal disorders,  
pregnancy-related disorders, endocrine disorders, and infections. The  
proteins can also be used to aid wound healing and epithelial cell  
proliferation, to prevent skin ageing due to sunburn, to maintain organs  
before transplantation, for supporting cell culture of primary tissues,  
to regenerate tissues, to identify their cognate ligands or binding  
partners, and in chemotaxis, and can be used as a food additive or  
preservative to modify storage properties. Antibodies specific for a  
protein of the invention can be used in alleviating symptoms associated  
with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
present sequence represents a human secreted protein-encoding cDNA of the  
invention

SQ Sequence 2615 BP; 616 A; 764 C; 717 G; 514 T; 0 U; 4 Other;

Alignment Scores:  
Pred. No.: 0 Length: 2615  
Score: 390.00 Matches: 731  
Percent Similarity: 98.92% Conservative: 0  
Best Local Similarity: 98.92% Mismatches: 5  
Query Match: 52.99% Indels: 8  
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x AAD08507 (1-2615)

QY 1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValValAlaAlaLaser 20  
DB 24 ATGGCGGTAGCGCGGTGAAGCTGTGACACACACTGCTGGCTGTCTGGCGCTGCTCC 83  
QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40  
DB 84 CAAGCCGAGTTCGAGTCCGAGGCGAGGATGGGCGATGGTGACGCTCATCTGCTCTCGCC 143  
QY 41 GluGlyThrAlaAlaValAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
DB 144 GAGGGGACCGACGCTACGCGCGCGGGACTGGCGGGGTGCTCTGAGCATGGAAACGG 203  
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
DB 204 CGCTGGCTCCCGGGGAGCGCTCCGCGCCCTTCGCTGGCTGGCGACCCAGTGTGCC 263  
QY 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
DB 264 GCCGACTTCGCGTGGGAGCTGGACCCGACTGTGCTCCCGACCGCGCGCTCGGGC 323  
QY 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLe 120  
DB 324 GCCGC-CGCCCTGGCGGACCTGAGCTTCTTCGGGGGCTTCTGCTGGCGCTCGCTGCC 382  
QY 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140  
DB 383 CGCGCGTGTCTGGCGCGCGCGCGCGC-CACTGCTCAGCGAAAGATGAGTGGAGTT 441  
QY 140 eArgLyArgSerProTrpAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGl 160  
DB 442 CCGCAAGCGGAC-CCCTACAACTACCTGCGAGTGGCTTCTCAAGATCAACAAAGTTGGA 500  
QY 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGl 180  
DB 501 GAAAGCTGTCTCGACGACACACACTTCTTCGTGGCAATCCTGAGCACATGGAATGCA 560  
QY 180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
DB 561 GCAGAACCTAGACTATTACCAAAACCATGTCTGAGAGTGAAGGCGGACACTCAAGGATCT 620  
QY 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGl 220  
DB 621 TGAGACTCAACCCCATATGCAAGAAATTCGACTGGAGTGGGACTCTACTCAGAGGACCA 680  
QY 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGl 240  
DB 681 GCCACAGAAAGCTGTGCCCCACCTAGAGCGGCGCTGCAAGATACATTGTGGCGCTATGA 740  
QY 240 uGluCysArgAlaLeuCysGlyGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTy 260  
DB 741 GGAGTGGCGTGGCTCTCGGAAGGCGCTATGACTACGATGGCTACAACTACCTTGAAGTA 800  
QY 260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGl 280  
DB 801 CAACGCTGACCTTCCAGGCGCATCACAGATCATTACATCCAGGTCTCTCAACTGTAAAGCA 860  
QY 280 rAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
DB 861 GAACTGTGTCCGAGCTGTCTCCACCCCAAGTCGAGAGGCGCTTGAAGACTTCCCT 920  
QY 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320

DB 921 CCCATGCAATTATTAATTATCTGCAGTTTGCCTACTATAACATTGGGAATTATACACAGC 980  
QY 320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetIleGlnA 340  
DB 981 TGT-TGAATGTGCCAAGACCTATCTTCTTCTTCCCAATGAGAGGTGATGAACAAA 1039  
QY 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG 360  
DB 1040 ATTGGCCCTATTATGACGTATGCTTGGAGAAGAACACACACAGATCCATCGGCCCGCTG 1099  
QY 360 luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaT 380  
DB 1100 AGAGTGCCCAAGAGTACCGACAGCAAGCCTACTTGGAAAAAGAACTGCTTTCTTCGCTT 1159  
QY 380 yrAspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIleP 400  
DB 1160 ATGATGTTTTTGGAAATCCCTTTTGGGATCCGATTCATGACATCCAGAAAGATGATTC 1219  
QY 400 rLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
DB 1220 CCAAGAGATTGCAAGAGAAACAGAAAGTCAGAAACGGGAAACAGCCGTACGCATCTCCCA 1279  
QY 420 luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440  
DB 1280 AGATTGGGAACCTTATGAAGAAATCGAGACCTTGTGGAAGAGAAAGAACAGGAGTCCAC 1339  
QY 440 euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuT 460  
DB 1340 TGGATGTGACGAGACTGACCCGGGAAGGTGGCCCTCTGTATGAGGCATCATCTCA 1399  
QY 460 hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer 479  
DB 1400 CCATGAACTCCAACTCCTGAATGGTTC-CCAGCCGGGTGGTGATGGAGCGGGAATCTCT 1458  
QY 480 AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThr-SerGlyAspGly 499  
DB 1459 GACCAAGAGTGTACGAGCTGCAGAGACTGACCATGTGGCAGCAACCTCAGAGATGSC 1518  
QY 500 TyrArgGlyGlnThr-SerProHisThrProAsnGluLysPheTyr-GlyValThrValPhe 519  
DB 1519 TACCGGGGTACAGCTCCCCACATCTCCCAATGAAAGTTCTATGGTGTCACTGTCTTC 1578  
QY 520 LysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyr 539  
DB 1579 AAAGCCCTCAAGCTGGGGCAAGAGCAAGTTCTCTGCAGAGTGCCACCTGACTACTAC 1638  
QY 540 AsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeu 559  
DB 1639 AACGTGACGGAGAGGTGGCGGCATCATGGAGTCTCTACTCCGCTGGATACGCCCTTC 1698  
QY 560 TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArg 579  
DB 1699 TACTTTTCTACTCTCATCTGTGTGGCGCACTGCCATCGAAGAGGTCCAGGACAGAGG 1758  
QY 580 LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599  
DB 1759 AAGATGATGATCATCCAGTCCAGTGGCAACTGCATCTCTGATGCCGAGACCTTCGTTG 1818  
QY 600 CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619  
DB 1819 TGTGTCAAAAGAGCCCCCAGCTACACTTCCGCGCATACAGCGCCATCTTACCTAAAT 1878  
QY 620 GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla 639  
DB 1879 GGGAATCTCGATGGCGGAACCTTTTATTTCATCTGAACTGGATGCCAAGACCGGTGACGCA 1938  
QY 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659  
DB 1939 GAGGTGACGCTCAGTGTGGAAGAGCGGTGGGATCTCTTCAGGCACTGAAACCCACAT 1998  
QY 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuAsp 679



Db 1999 GGAGTGAAGGCTGTACCAAGGGGAGCGCTGTGCATCGCCCTGTGTTACCCCTGGAC 2058  
 QY 680 ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValValMetLeuPheSer 699  
 Db 2059 CTTGACACAGGAGCGGACAGGGTGCAGGAGATGACCTGGTGAAGATCTCTTCAGC 2118  
 QY 700 ProGluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGlu 719  
 Db 2119 CCAGAAGAGATGAGCTCTCCAGGAGCGCCCTGGATGCCAGCAGGGCCCCCGGAA 2178  
 QY 720 ProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
 Db 2179 CCTGCACAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2229

## RESULT 9

ABK94910

ID ABK94910 standard; cDNA; 2753 BP.

XX AC ABK94910;

XX DT 30-AUG-2002 (first entry)

XX DE Human novel polynucleotide #21.

XX KW Human; gene; ss; inflammatory condition; shock; sepsis; immune response;

XX KW cancer; wound healing; central nervous system disease; haematopoiesis;

XX KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;

XX KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;

XX KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;

XX KW bone degenerative disorder; periodontal disease; reperfusion injury;

XX KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;

XX KW allergic condition; thrombolytic; thrombosis; coagulation disorder;

XX KW fungal infection.

XX OS Homo sapiens.

XX PN WO200244340-A2.

XX PD 06-JUN-2002.

XX PF 30-NOV-2001; 2001WO-US047004.

XX PR 30-NOV-2000; 2000US-00728952.

XX PS (HYSE-) HYSEQ INC.

XX PI Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;

XX PI Yamazaki V, Ujwal ML, Drmanac RT;

XX DR WPI; 2002-508509/54.

XX DR P-PSDB; ABG66686.

XX PT Novel nucleic acids and polypeptides for diagnosis, treatment of

XX PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell

XX PT disorders, cancer and promoting wound healing.

XX PS Claim 1; Page 369-373; 672pp; English.

XX CC The invention relates to human novel polynucleotides and associated

XX CC polypeptides. The polynucleotides and polypeptides are useful for

XX CC treating inflammatory conditions such as arthritis, nephritis, Crohn's

XX CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses

XX CC and cancer and promoting wound healing. The sequences are used to

XX CC induce the proliferation of neural cells and regeneration of nerve and

XX CC brain tissue, and are useful for the treatment of central and peripheral

XX CC nervous system diseases and neuropathies, such as Alzheimer's disease,

XX CC Parkinson's disease, Huntington's disease and amyotrophic lateral

XX CC sclerosis. The sequences are involved in chemotactic or chemokinetic

XX CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid

XX CC cell disorders and platelet disorders such as thrombocytopenia,

XX CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue

XX CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of

XX CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal

XX CC

CC disease. The sequences of the invention are also useful for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues, immune deficiencies and disorders  
 CC including severe combined immunodeficiency (SCID), bacterial or fungal  
 CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia  
 CC gravis, allergic conditions such as asthma, thrombolytic or thrombotic  
 CC and coagulation disorders. Sequences ABK94890-ABK94982 represent human  
 CC novel polynucleotides of the invention  
 XX

SQ Sequence 2753 BP; 632 A; 802 C; 784 G; 535 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 0 Length: 2753  
 Score: 380.00 Matches: 580  
 Percent Similarity: 99.32% Conservative: 0  
 Best Local Similarity: 99.32% Mismatches: 2  
 Query Match: 51.63% Indels: 4  
 DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x ABK94910 (1-2753)

QY 155 LysIleAsnLysLeuGlnLysAlaValAlaAlaHisThrPhePheValGlyAsnPro 174  
 Db 673 AAGATCAACAAAGTTGAGAAAGCTGTGTGCAGCACACACCTCTTCTGTGGCAATCT 732  
 QY 175 GluHisMetGluMetGlnGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGlu 194  
 Db 733 GAGCATGTGAAATGCAGCAGCACTAGCTATTACCAACCATCTCTGGAGTGAAGAG 792  
 QY 195 AlaAspPheLysAspLeuGlnThrGlnProHisMetGlnGluPheArgLeuGlyValArg 214  
 Db 793 GCGCATCTCAGGATCTTGAGACTCAACCCCATATGCAAGAAATTCGACTGGAGTGGCA 852  
 QY 215 LeuTyrSerGluGluGlnProGlnAlaValProHisLeuGluAlaAlaLeuGlnGlu 234  
 Db 853 CTCTACTCAGAGAAACAGCCAGGAAGCTGTGCCCCACCTAGAGCGGCGCTGCAAGAA 912  
 QY 235 TyrPheValAlaTyrGluGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGly 254  
 Db 913 TACTTTGTGGCTATGAGAGTGGCTGCTCTCGAAGGGCCCTATGACTAGATGCG 972  
 QY 255 TyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGln 274  
 Db 973 TACAACCTACCTTGATGATCAACGCTGACCTCTCCAGGCGCATCACAGATCATATCATCCAG 1032  
 QY 275 ValLeuAsnCysLysGlnAsnCysValThrGluLeuAlaSerHisProSerArgGluLys 294  
 Db 1033 GTCTCAACTGTAGCAGCAACTGTGTACAGGAGCTTGTCTCCACCCCAAGTCGAGAGAG 1092  
 QY 295 ProPheGluAspPheLeuProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIle 314  
 Db 1093 CCCTTTGAGACTTCTCCCATGCAATTAATATATCTGCAGTTTGGCTACTATAACATT 1152  
 QY 315 GlyAsnTyrThrGlnAlaGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAs 334  
 Db 1153 GGGATTTATACAGGCTGT-TGAATGTGCCAAGACCTATCTTCTCTTCTTCCCAATGA 1211  
 QY 334 pGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGluHisThrAr 354  
 Db 1212 CGAGTGTATGAACCAAAATTTGGCTATTATGAGCTATGCTTGGAGAGAAACACACAG 1271  
 QY 354 gSerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeuLeuGlyLysG 374  
 Db 1272 ATCATCGGCCCCCGTGAGAGTGCAGAGGAGTCCGACGAGGAGGCTACTTGGAAAAAGA 1331  
 QY 374 uLeuLeuPhePheAlaTyrAspValPheGlyIleProPheValAspProAspSerTrpTh 394  
 Db 1332 ACTGCTTTCTTCGCTTATGATGTTTGGAAATTCCTTTGTGGATCCGATTCATGGAC 1391  
 QY 394 rProGluGluValIleProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAl 414  
 Db 1392 TCCAGAGAGAGTGAATCCCAAGAGATTGCAAGAGAAACAGAAAGTCAGCGGAAAAAGCAG 1451



QY	414	aValArgIleSerGlnGluLeuGlyAsnLeuMetLysGluLeuThrLeuValGluGI	434
Db	1452	CGTACGCATCTCCAGAGATTGGGAACCTTATGAAGGAATCGAGACCCCTTGTGAAGA	1511
QY	434	uLysThrLysGluSerLeuAspValSerArgLeuThrArgGluGluGlyProLeuLeuTy	454
Db	1512	GAAGACCAAGAGTCACTGATGTGACGACACTGACCGGAAGTGGCCCTGTCTGTA	1571
QY	454	rGluGlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValM	474
Db	1572	TGAAGGCATCAGTCTCACCATGAATCAACCTCAAGTGGTTC-CCAGCGGTGGTGA	1630
QY	474	eAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaA	494
Db	1631	TGGACGCGCTAATCTTGACCAAGAGTGTGAGAGCTGCAGAGACTGACCAATGTGCAG	1690
QY	494	laThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheT	514
Db	1691	CAACCTCAGGAGATGGTACGGGCTCAGACCTCCCCACATACTCCCAATGAAGAATCT	1750
QY	514	YrGlyValThrValPheIleAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnS	534
Db	1751	ATGGTGTCACCTGCTTCAAGCCCTCAAGCTGGGGCAAGGAAGTTCCTCTGCAGA	1810
QY	534	erAlaHisLeuTyrTyrAsnValThrGluLysValArgArgIleMetGluSerTyrPheA	554
Db	1811	GTGCCCACTGTACTACACGTGACGGAAGTGGCGGCATCATGGAGTCTCTCTTC	1870
QY	554	rgLeuAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluG	574
Db	1871	GCCTGGATACGCCCTCTACTTTCTCTACTCTCATCTGGTGTGCCGCACTGCCATGAAG	1930
QY	574	luValGlnAlaGluArgLysAspSerHisProValHisValAspAsnCysIleLeuA	594
Db	1931	AGGTCAGGCGAGAGAGATGATGATCATCGTCCAGTCCAGTCACTGCATCTCTGA	1990
QY	594	snAlaGluThrLeuValCysValLysGluProProAlaTyrThrPheArgAspTyrSerA	614
Db	1991	ATGCCGAGACCTCGTGTGTGTCAAGAGCCCCAGCCCTACACCTTCCGCGACTACAGCG	2050
QY	614	laIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspA	634
Db	2051	CCATCCTTTACTTAATGGGAGCTTCGATGGCGGAACCTTTATTTCACTGAACTGGATG	2110
QY	634	laLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPheSerSerG	654
Db	2111	CCAAGCCGTACGCGCAGAGCTGACGCTCAGTGTGGAAGAGCCGTGGGATTCTCTTCAG	2170
QY	654	lyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaL	674
Db	2171	GCACTGAAACCCACATGGAGTGAAGGCTGTACACAGGGGGCAGCGTGTGCCATCGCCC	2230
QY	674	euTpPheThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspAspLeuV	694
Db	2231	TGTGTTTCACTTGACCTCGACACAGCAGCGGACAGGGTGCAGGCAGATGACCTGG	2290
QY	694	alLysMetLeuPheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspIaG	714
Db	2291	TGAAGATCTCTTCAGCCCGAAGAGATGGACCTCTCCAGAGACACCCCTGGATGCC	2350
QY	714	lnGlnGlyProProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysA	734
Db	2351	AGCAGGGCCCCCGAACCTGCACAGAGTCTCTCTCAGGCAGTGAATCGNAGCCCAAGG	2410
QY	734	spGluLeu	736
Db	2411	ATGAGCTA	2418
XX	RESULT 10		
XX	ABK94933		
XX	ID	ABK94933	standard; cDNA; 2753 BP.
XX	AC	ABK94933;	

XX	30-AUG-2002	(first entry)	
DT	Human novel polynucleotide #44.		
DE	Human; gene; ss; inflammatory condition; shock; sepsis; immune response;		
XX	cancer; wound healing; central nervous system disease; haematopoiesis;		
KW	peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;		
KW	myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;		
KW	carilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;		
KW	bone degenerative disorder; periodontal disease; reperfusion injury;		
KW	lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;		
KW	allergic condition; thrombolysis; thrombosis; coagulation disorder;		
XX	fungal infection.		
OS	Homo sapiens.		
XX	WO200244340-A2.		
PN	06-JUN-2002.		
XX	30-NOV-2001; 2001WO-US047004.		
XX	30-NOV-2000; 2000US-00728952.		
PR	(HYSE-) HYSEQ INC.		
XX	Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;		
PA	Yamazaki V, Ujwal ML, Drmanac RT;		
PI	WPI; 2002-508509/54.		
XX	P-PSDB; ABG66709.		
DR	Novel nucleic acids and polypeptides for diagnosis, treatment of		
XX	inflammatory, autoimmune, nervous system, myeloid or lymphoid cell		
PT	disorders, cancer and promoting wound healing.		
XX	Claim 1; Page 441-444; 672pp; English.		
PS	The invention relates to human novel polynucleotides and associated		
XX	polypeptides. The polynucleotides and polypeptides are useful for		
CC	treating inflammatory conditions such as arthritis, nephritis, Crohn's		
CC	disease, ischaemia-reperfusion injury, shock, sepsis, immune responses		
CC	and cancer and for promoting wound healing. The sequences are used to		
CC	induce the proliferation of neural cells and regeneration of nerve and		
CC	brain tissue, and are useful for the treatment of central and peripheral		
CC	nervous system diseases and neuropathies, such as Alzheimer's disease,		
CC	Parkinson's disease, Huntington's disease and amyotrophic lateral		
CC	sclerosis. The sequences are involved in chemotactic or chemokinetic		
CC	activity, regulation of haematopoiesis, treatment of myeloid or lymphoid		
CC	cell disorders and platelet disorders such as thrombocytopenia, tissue		
CC	regeneration of bone, cartilage, tendon, ligament and/or nerve tissue		
CC	growth, tissue repair, healing of burns, incisions, ulcers, treatment of		
CC	osteoporosis, osteoarthritis, bone degenerative disorders and periodontal		
CC	disease. The sequences of the invention are also useful for gut		
CC	protection or regeneration and treatment of lung or liver fibrosis,		
CC	reperfusion injury in various tissues, immune deficiencies and disorders		
CC	including severe combined immunodeficiency (SCID), bacterial or fungal		
CC	infections, autoimmune disorders e.g. multiple sclerosis and myasthenia		
CC	gravis, allergic conditions such as asthma, thrombolysis or thrombosis		
CC	and coagulation disorders. Sequences ABK94890-ABK94982 represent human		
CC	novel polynucleotides of the invention		
XX	Sequence 2753 BP; 632 A; 802 C; 784 G; 535 T; 0 U; 0 Other;		
SQ	Alignment Scores:		
	Pred. No.:	0	Length: 2753
	Score:	380.00	Matches: 580
	Percent Similarity:	99.32%	Conservative: 0
	Best Local Similarity:	99.32%	Mismatches: 2
	Query Match:	51.63%	Indels: 4
	DB:	6	Gaps: 0

US-10-045-815-4 (1-736) x ABK94933 (1-2753)

QY 155 LysIleAsnLysLeuGluLysAlaValAlaAlaAlaHisThrPheValGlyAsnPro 174  
Db 673 AAGATCAACAAGTTCGGAAGAGCTGTCTGCTGAGCAGACACCTTCTCGTGGCAATCCT 732  
QY 175 GluHisMetGluMetGlnGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGlu 194  
Db 733 GAGCACATGGAAATGCACAGAACCTAGACTATTACCAACCATGTCTGGAGTGAAGGAG 792  
QY 195 AlaAspPheLysAspLeuGluThrGlnProHisMetGlnGluPheArgLeuGlyValArg 214  
Db 793 GCCGACTTCAAGGATCTTGAAGACTCAACCCCATATGCAAGAAATTCGAGTGGAGTGGCA 852  
QY 215 LeuTyrSerGluGluGlnProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGlu 234  
Db 853 CTCTACTCAGAGGAACACCCACAGAGCTGTGCCACCTAGAGGGCGGCTGCAAGAA 912  
QY 235 TyrPheValAlaTyrGluGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGly 254  
Db 913 TACTTTGTGGCTATGAGGAGTGGCTGCTCTGCGAAGGGCCCTATGACTACGATGGC 972  
QY 255 TyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGln 274  
Db 973 TACAACCTACCTTGAGTACACGCTGACCTCTTCCAGGCCATCAGATCATATTACATCCAG 1032  
QY 275 ValLeuAsnCysLysGlnAsnCysValThrGluLeuAlaSerHisProSerArgGluLys 294  
Db 1033 GTCCCTCACTGAGCAGCACTGTGTACGAGCTTGTCTCCCAACCAAGTCGAGAGAAG 1092  
QY 295 ProPheGluAspPheLeuProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIle 314  
Db 1093 CCTTTTGAAGACTTCTCTCCCATCGCAATATATATCTGCGAGTTGGCTTACTATAACATT 1152  
QY 315 GlyAsnTyrThrGlnAlaGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAs 334  
Db 1153 GGGAAATTATACAGAGCTGT-TGAATGTGCCAGACCTATCTTCTTCTTCCCAATGA 1311  
QY 334 pGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrAr 354  
Db 1212 CGAGGTGATGAACCAAAATTTGGCTATTATGAGCTATGCTTGGAGAAACACACACAG 1271  
QY 354 sSerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysG 374  
Db 1272 ATCCATCGCCCCCGTGAGAGTGCAGAGGATTCGAGACGAGGAGCTACTGGAAAAAGA 1331  
QY 374 uLeuLeuPhePheAlaTyrAspValPheGlyIleProPheValAspProAspSerTrpTh 394  
Db 1332 ACTGCTTTCTTCGCTTATGATGTTTTTGGAAATTCCTTTGTGGATCCGGATTTCATGGAC 1391  
QY 394 rProGluGluValIleProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAl 414  
Db 1392 TCCAGAAAGAGTGAATCCCAAGAGATTGCAAGAGAAACGAGGTCAGAACCGGAAACAGC 1451  
QY 414 aValArgIleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThrLeuValGluG 434  
Db 1452 CGTACGCATCTCCAGAGAGTGGGAACCTTATGAGGAAATCGAGACCTCTTGGAGA 1511  
QY 434 uLysThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTy 454  
Db 1512 GAAGACCAAGGAGTCACTGGATGTGAGCAGACTGACCCGGGAAGGTGGCCCCCTGTCTGA 1571  
QY 454 rGluGlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValM 474  
Db 1572 TGAAGGATCATGCTCCACCATGAACTCCAACTCCTGAAATGGTTC-CCAGCGGGTGGTGA 1630  
QY 474 etAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAla 494  
Db 1631 TGGACGGCGTAATCTCTGACCAACAGGTGTGAGGAGCTGACGAGACTGACCAATGTGGCAG 1690  
QY 494 laThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheT 514  
Db 1691 CAACCTCAGAGATGGCTACCGGGGTGAGACCTCCCAACATATCCCAATGAAAATTTCT 1750

QY 514 YrClyValThrValPheLysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnS 534  
Db 1751 ATGGTGTCACTGTCTTCAAGGCCCTCAAGCTGGGCAAGAGGCAAGGTTCCTCTGCAGA 1810  
QY 534 erAlaHisLeuTyrTyrAsnValThrGluLysValArgArgIleMetGluSerTyrPheA 554  
Db 1811 GTGCCACCTGTACTACAACGTGACGGAGAAGGTGGCGCATCATGAGTCTTACTTCC 1870  
QY 554 rGluAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluG 574  
Db 1871 GCTGTGATAGCCCTCTACTTTTCTACTCTCATCTGTGTGCGGCACATGCGCATCGAAG 1930  
QY 574 luValGlnAlaGluArgLysAspSerHisProValHisValAspAsnCysIleLeuA 594  
Db 1931 AGGTCCAGCAGAGAGGAAGATGATGTCATCCAGTCCACGTGGACCACTGTCATCTGA 1990  
QY 594 snAlaGluThrLeuValCysValLysGluProProAlaTyrThrPheArgAspTyrSerA 614  
Db 1991 ATGCCGAGACCTCTGTGTGTGTCAAAGAGCCCCAGCCTACACCTTCGCGACTACAGC 2050  
QY 614 laIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspA 634  
Db 2051 CCATCTTTTACCTAAATGGGACTTTCGATGGCGGAAACTTTTATTTCACTGAACCTGGAT 2110  
QY 634 laLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPheSerSerG 654  
Db 2111 CCAAGACCGTGACGCGCAGAGGTGCGAGCTCAGTGTGGAAGAGCGGTGGATTCCTTCAG 2170  
QY 654 lyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaL 674  
Db 2171 GCATCGAAACCCACATGAGTGAAGGTGTCCACGAGGGCGAGCGCTGTGCCATCGCCC 2230  
QY 674 euTyrPheThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspAspLeuV 694  
Db 2231 TGTGTTTCACTTGGACCTCGACACAGCGAGCGGACAGGGTGCAGGCGATGACCTGG 2290  
QY 694 aLysMetLeuPheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaG 714  
Db 2291 TGAAGATGCTCTTTCAGCCAGAGAGATGAGACCTCTCCAGAGAGCAGCCCTGTGATGCC 2350  
QY 714 lngGlnGlyProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysA 734  
Db 2351 AGCAGGGCCCCCCCCAACCTGCACAAAGATCTCTCTCAGGCGATGATCGAAGCCCAAG 2410  
QY 734 spGluLeu 736  
Db 2411 ATGAGCTA 2418

## RESULT 11

AAC64724  
ID AAC64724 standard; cDNA; 2829 BP.

XX AAC64724;

XX 27-FEB-2001 (first entry)

DE Human tumour suppressor Gros1-L encoding cDNA SEQ ID NO:1.

XX Tumour suppressor; Gros1-L; Gros1-S; cell proliferation; regulation;  
KW cancer; cytostatic; gene therapy; ss.

XX Homo sapiens.

XX WO200065047-A1.

XX 02-NOV-2000.

XX 26-APR-2000; 2000WO-JP002731.

XX 26-APR-1999; 95JP-00118806.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX Wadhwa R, Sugihara T, Yoshida A;  
XX WPI: 2000-687340/67.  
XX P-PSDB; AAB36391.  
XX Full-length tumor suppressor genes encoding Grosi-L, Grosi-S and mouse  
PT homologs participating in regulation of cell proliferation, useful in  
PT development of preventives and remedies of cancer.  
XX Claim 1; Page 61-66; 114pp; Japanese.  
XX The present sequence encodes the human tumour suppressor designated Grosi  
CC -L, Grosi-L and Grosi-S have cytosolic activity and can be used in gene  
CC therapy. Grosi-L and Grosi-S genes are useful in the development of drugs  
CC used to treat and prevent cancer  
XX Sequence 2829 BP; 659 A; 808 C; 801 G; 561 T; 0 U; 0 Other;  
SQ  
Alignment Scores:  
Pred. No.: 0 Length: 2829  
Score: 360.00 Matches: 360  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 48.91% Indels: 0  
DB: 3 Gaps: 0  
US-10-045-815-4 (1-736) x AAC64724 (1-2829)  
QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaLaser 20  
DB 52 ATGGCGGTACGCGGTTGAAGTGTGACACACTGCTGGCTGTGGCGCGCTGCTGC 111  
QY 21 GlnAlaGluValGluSerGluAlaGlyTTPGlyMetValThrProAspLeuPheAla 40  
DB 112 CAAGCCGAGGTGAGTCCGAGGAGGATGGGATGGTGTGACGCTGATCTGCTCTGCC 171  
QY 41 GluGlyThrAlaAlaValArgGlyAspTTPProGlyValValLeuSerMetGluArg 60  
DB 172 GAGGGACCCGAGCTACGCGCGCGGAGCTGGCCGGGTGTCTCGATGAGTGAACGG 231  
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
DB 232 GCGTGGCTCCGGGCGAGCCCTCCGCGCCCTTCGCTGGCTGGCGCACCCAGTGTGC 291  
QY 81 AlaAspPheProTTPGluLeuAspProAspTTPSerProAlaGlnAlaSerGly 100  
DB 292 GCCGACTTCCGCTGGGAGCTGGACCCGACTGTCTCCCGAGCCGCGCCCTCGGC 351  
QY 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyLeuLeuArgArgAlaAlaCysLeu 120  
DB 352 GCGCGCCCTCGCGACCTGAGCTTCTTCGGGGGCTTCTGCGTGGCTGCTGCTGCTG 411  
QY 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluLeuMetGluLeuPhe 140  
DB 412 CGCCGCTGCTCGGGCGCGCGCCGCTGCTGCTGAGGAGAGTGGAGTTC 471  
QY 141 ArgLysArgSerProTTPAsnTyrLeuGlnValAlaTyrPheLysLeuLeuGlu 160  
DB 472 CGCAAGCGGAGCCCTACACTACTTGCAGTGTGCTTCAAGATCAACAAGTTGGAG 531  
QY 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180  
DB 532 AAAGCTGTGTGTCAGCACACACCTTCTTGTGGGCAATCTGAGCACATGGAATGCAG 591  
QY 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
DB 592 CAGAACCTAGACTATTACCAACCATCTCTGGAGTGAAGGCGCCGACTTCAGAGATCTT 651  
QY 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220  
DB 652 GAGACTCAACCCCATATGCAAGATTTTCGACTGGAGTGGGACTCTACTCAGAGGAACAG 711

QY 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
DB 712 CCACAGGAAGCTGTGCCCCACCTAGAGCGCGCTGCAAGAAATACTTTTGTGGCCTATGAG 771  
QY 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
DB 772 GAGTCCCGTGGCCCTCTGCAAGGCGCTTATGACTACGATGCTACACTACTCTTGAGTAC 831  
QY 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
DB 832 AACGCTGACCTTCTCCAGGCCATCACAGATCATCATCCAGGCTCCTCACTGTAAGCAG 891  
QY 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
DB 892 AACTGTGTACAGAGCTTGTCTCCCAAGTCAGAGAAGCCCTTTTGAAGACTTCTTC 951  
QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
DB 952 CCATCGCATTAATTAATCTGAGTTTGCCTTACTATAACATTGGGATTTATACACAGCT 1011  
QY 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
DB 1012 GGTGAATGTGCAAGACCTATCTTCTTCTTCCCAATGACGAGGTGATGAACCAAAAT 1071  
QY 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
DB 1072 TTGGCCTATTATGACGCTATGCTTGGAGAAGACACACAGATCCATCGCCCGCTGAG 1131  
RESULT 12  
AAH16288  
ID AAH16288 standard; cDNA; 2993 BP.  
XX AAH16288;  
XX 26-JUN-2001 (first entry)  
XX Human cDNA sequence SEQ ID NO:15159.  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
XX EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-00116126.  
XX 29-JUL-1999; 99JP-00248036.  
XX 27-AUG-1999; 99JP-00300253.  
XX 11-JAN-2000; 2000JP-00118776.  
XX 02-MAY-2000; 2000JP-00183767.  
XX 09-JUN-2000; 2000JP-00241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI: 2001-318749/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
PT length cDNAs defined in the specification, and for the detection and/or  
PT diagnosis of the abnormality of the proteins encoded by the full-length  
PT cDNAs.  
XX Claim 8; SEQ ID NO 15159; 2537pp + Sequence Listing; English.  
XX The present invention describes primer sets for synthesising 5602 full-  
CC length cDNAs defined in the specification. Where a primer set comprises:  
CC (a) an oligo-dr primer and an oligonucleotide complementary to the  
CC complementary strand of a polynucleotide which comprises one of the 5602  
CC nucleotide sequences defined in the specification, where the

oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 2993 BP; 666 A; 864 C; 856 G; 607 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 1.33e-282 Length: 2993  
Score: 295.00 Matches: 681  
Percent Similarity: 98.84% Conservatives: 0  
Best Local Similarity: 98.84% Mismatches: 4  
Query Match: 40.08% Indels: 8  
DB: 4 Gaps: 0

US-10-045-815-4 (1-736) x AAH16288 (1-2993)

Qy 1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaSer 20  
Db 42 ATGGCGGTACCGCGGTGAACCTGCTGACCACTGCTGGCTGCGGGCGCTGCTCC 101  
Qy 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40  
Db 102 CAAGCCGAGGTCCGAGTCCGAGGAGGATGGGCATGTGTGACGCTGATCTGCTTCGCC 161  
Qy 41 GluGlyThrAlaAlaValArgAlaGlyAspTrpProGlyValValLeuSerMetGluArg 60  
Db 162 GAGGGGACCGGAGCTTACCGCGCGGGAGCTGGCCCGGGTGTCTTGGACATGGAAACGG 251  
Qy 61 AlaLeuArgSerArgAlaAlaLeuAlaLeuArgAlaLeuArgCysArgThrGlnCysAla 80  
Db 222 GCGCTGCGCTCCCGGCGAGCCCTCCGCGCCCTGCGCTGCGCTGCCACCCAGTGTGCC 281  
Qy 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
Db 282 GCCGACTTCCCGTGGGAGCTGGACCCGACTGGTTCGCCCGCGGCCAGGCTCGGGC 341  
Qy 101 AlaGly-AlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgArgAlaAlaCysLe 120  
Db 342 GCCGC-CCGCTGCGGACCTGAGCTTCTCGGGGGCTTCTGCGTGGCTGCGCTGCGCT 400  
Qy 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe 140  
Db 401 GCGCCGCTGCTCGGGCGCGCGCGCCCTCGCTCAGGAAAGAGATGAGGTGGAGT 460  
Qy 140 eArgLysArgSerProTrpAsnTrpLeuGlnValAlaTyrPheLysIleAsnLysLeuG 160  
Db 461 CCGCAACGGAGCCCTTACAACTACCTGAGGTGCGCTGCTTCAAGATCAACAAAGTTGGA 520  
Qy 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetG 180  
Db 521 GAAAGCTGTTGCTCGAGCACACACTTCTTCGTGGGCAATCCTGAGCACATGGAATGCA 580  
Qy 180 nGlnAsnLeuAspTrpTrpGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
Db 581 GCAGAACTAGACTATTACCAACCACTGCTGGAGTGAAGAGGCCGACTTCAAGGATCT 640  
Qy 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTrpSerGluGlu 220  
Db 641 TGAGACTCAACCCCATATGCAAGAAATTTTCAGCTGGGAGTCCGACTTCTACTCAGAGGA 700

Qy 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTrpPheValAlaTrpG 240  
Db 701 GCCACAGGAGCTGTGCCCCACCTAGAGCGGCGCTGCAAGATACCTTTGTGGCCTATGA 760  
Qy 240 uGluCysArgAlaLeuCysGluGlyProTrpAspTrpAspGlyTrpAsnTrpLeuGluTr 260  
Db 761 GGAGTGGCGTGGCCCTCTCGAAGGCGCTATGACTACCATGGCTACACTACCTTGAGTA 820  
Qy 260 rAsnAlaAspLeuPheGlnAlaLeuThrAspHisTrpIleGlnValLeuAsnCysLysG 280  
Db 821 CAACGCTGACCTTCTCCAGGCCATCAGATCATTCATCCAGGTCTCACTGTAAGCA 880  
Qy 280 nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
Db 881 GAACTGTGTACGAGCTTGTCTCCACCCCAAGTCAGAGAAGCCCTTTGAAGACTTCTCT 940  
Qy 300 uProSerHisTrpAsnTrpLeuGlnPheAlaTrpTrpAsnIleGlyAsnTrpThrGlnAl 320  
Db 941 CCCATCGCATTATAATTATCTGAGTTTGCCTTCTTCCCTATTAACATTGGGAATTATCACAGGC 1000  
Qy 320 agly-GluCysAlaLysThrTrpLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
Db 1001 TGT-TGAATGTGCCAAGACCTATCTTCTTCTTCCCAATGACGAGGTGATGAACAAA 1059  
Qy 340 snLeuAlaTrpTrpAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG 360  
Db 1060 ATTTGGCCTTATTATGAGCTATGCTTGGAGAAGAACACACAGATCCATCCGCCCGCTG 1119  
Qy 360 LuSerAlaLysGluTrpArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaTr 380  
Db 1120 AGAGTGCACAGAGTACCGACAGGAGCTTCTTGGAAAAGAACTGCTTTTCTCGCTT 1179  
Qy 380 yAspValPheGlyIleProPheValAspProAspSerTrpThrProGlu-GluValIle 399  
Db 1180 ATGATGTTTTGGAAATTCCTTTGTGGATCCGGATTCTATGGACTCCAGG-AGAAAGTGAT 1238  
Qy 400 ProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGln 419  
Db 1239 CCCAAGAGATTGCAAGAGAAACAGAGTCAAGACGGGAAACAGCCGTACGATCTCCAG 1298  
Qy 420 GluIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSer 439  
Db 1299 GAGATTGGGAACCTTATGAAGAAATCGAGACCTTGTGGAGAAGAGAACCAAGAGTCA 1358  
Qy 440 LeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTrpGluGlyLysSerLeu 459  
Db 1359 CTGGATGTGAGCAGACTGACCCGGAGGTGGCCCTCTGTATGAAGGATCAGTCTC 1418  
Qy 460 ThrMetAsnSerLysLeuLeuAsnGlyTrp-GlnArgValValMetAspGlyValIleSe 479  
Db 1419 ACCATGAATCCAAACTCCTGAATGGTTC-CCAGCGGTGGTGTGAGCGCGTAATCTC 1477  
Qy 479 rAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspG 499  
Db 1478 TGACCACAGTGTACAGGAGCTGACAGACTGACCAATGTGGCAGCAACCTCAGAGATGG 1537  
Qy 499 yTrpArgGlyGlnThrSerProHisThrProAsnGluLysPheTrpGlyValThrValPh 519  
Db 1538 CTACCGGGTTCAGACTTCCCCACATCTCCCAATGAAAAGTTCTATGGTGTCTGCTT 1597  
Qy 519 elysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTrpTy 539  
Db 1598 CAAGACCTCAAGCTGGGGCAAGGCAAGTTCCTCTGAGAGTCCCACTGCTGTACTA 1657  
Qy 539 rAsnValThrGluLysValArgArgIleMetGluSerTrpPheArgLeuAspThrProLe 559  
Db 1658 CAACGTCAGCGAGAGGTGCGCGCATCATGAGTCTTCTTCCGCTGGATACGCCCT 1717  
Qy 559 uTrpPheSerTrpSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluAr 579  
Db 1718 CTACTTTTCTACTCTCATCTGCTGTCGCGCACTGCCATCGAAGAGGTCCAGGAGAG 1777

QY 579 gLyAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVa 599  
 Db 1778 GAAGGATGATGATCATCCAGTCCACGTGGACAACTGCATCCTGAATGCCGAGACCTCGT 1837  
 QY 599 lCysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAs 619  
 Db 1838 GTGTGTCAAGAGCCGCCAGCTACACCTTCGCGACTACAGGCCATCCTTTACCTAAA 1897  
 QY 619 nGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAl 639  
 Db 1898 TGGGGACTTCGATGGCGGAACATTTTATTTCACTGAACCTGGATGCCCAACCGGTGACGCG 1957  
 QY 639 aGluValGlnProGlnCysGlyArgAlaValAlGlyPheSerSerGlyThrGluAsnProHi 659  
 Db 1958 AGAGGTGCACCTCAGTGTGGAGAGCCGTGGGATTTCTCTTCAGGCACCTGAAACCCACA 2017  
 QY 659 sGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAs 679  
 Db 2018 TGGAGTGAAGCGTGTCCACAGGGCGGAGCGCTGCGCATGCGCCCTGTGTTCCACCTGGA 2077  
 QY 679 pProArgHisSerGluArg 685  
 Db 2078 CCCTCGACACAGCGAGCGG 2096

## RESULT 13

AA517572  
 ID AA517572 standard; cDNA; 2127 BP.  
 AC AA517572;  
 XX 26-FEB-2002 (first entry)  
 DT DNA encoding novel secreted protein #1.  
 DE  
 XX Secreted protein; cytostatic; immunosuppressive; vulnery; vaccine;  
 KW antiinflammatory; neuroprotective; nephrotropic; cardiovascular; human;  
 KW cancer; autoimmune disease; wound healing disorder; infection;  
 KW haematopoietic disorder; inflammatory disorder; infertility;  
 KW neurological disease; psychiatric disease; cardiovascular disease;  
 KW respiratory disease; renal; gastrointestinal; ss.  
 XX Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 CDS 1..2127  
 FT /\*tag= a  
 FT /product= "Human secreted protein"  
 XX  
 PN WO200179454-A1.  
 XX  
 PD 25-OCT-2001.  
 XX  
 XX 11-APR-2001; 2001WO-US011797.  
 XX  
 XX 13-APR-2000; 2000US-0196603P.  
 XX 24-APR-2000; 2000US-0199417P.  
 XX  
 XX (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX  
 XX Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z;  
 XX WPI; 2002-061975/08.  
 DR P-PSDB; AAU09860.  
 XX  
 XX New secreted proteins or polypeptides, useful for treating e.g. cancer,  
 PT autoimmune diseases, wound healing disorder, infections, hematopoietic  
 PT disorders, inflammatory disorders, infertility, cancer.  
 XX  
 PS Claim 2; Page 34-35; 92pp; English.  
 XX  
 XX The invention relates to an isolated novel secreted polypeptide (I) and  
 CC polynucleotide (II). (I) and (II) are useful for treating cancer,

CC autoimmune diseases, wound healing disorder, infections, haematopoietic  
 CC disorders, inflammatory disorders, infertility, neurological and  
 CC psychiatric diseases, cardiovascular diseases, respiratory diseases,  
 CC renal diseases, or gastrointestinal diseases. These may also be used to  
 CC treat diseases, abnormalities and disorders caused by abnormal  
 CC expression, production, function and/or metabolism of the genes, as  
 CC vaccines for inducing immunological response in a mammal, and in  
 CC screening methods for detecting the effect of added compounds on the  
 CC production of mRNA and polypeptide in cells. The polypeptides can be used  
 CC as immunogens to produce antibodies immunospecific for the polypeptides,  
 CC and to identify membrane-bound or soluble receptors. The polynucleotides,  
 CC may be used as diagnostic reagents, in chromosome localisation studies,  
 CC and in tissue expression studies. The present sequence represents the  
 CC coding sequence of novel human secreted protein #1

SQ Sequence 2127 BP; 495 A; 623 C; 589 G; 420 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 7.31e-263 Length: 2127  
 Score: 275.00 Matches: 375  
 Percent Similarity: 99.47% Conservative: 0  
 Best Local Similarity: 99.47% Mismatches: 1  
 Query Match: 27.36% Indels: 2  
 DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x AA517572 (1-2127)

QY 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhePheAlaTyr 380  
 Db 997 AGTGCACAGAGTACCGACAGCGAGCTACTGGAAAAGAACTCTTTCTTCGCTAT 1056  
 QY 381 AspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIlePro 400  
 Db 1057 CATGTTTTTGGAAATTCCTTTTGTGATCCGGATTTCATGCACTCCAGAGAAAGTATTC 1116  
 QY 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
 Db 1117 AAGAGATTGCAAGAGAAACAGAGTACAGACGGAAACAGCGGTACGGATTCTCCAGAG 1176  
 QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
 Db 1177 ATTGGAACTTATGAAGGAAATCGAGACCTTGTGGAAAGAGAGACCAAGAGTCACTG 1236  
 QY 441 AspValSerArgLeuThrArgGluGlyClyProLeuLeuTyrGluGlyIleSerLeuThr 460  
 Db 1237 GATGTGACAGACTGACCCGGGAAGGTGGCCCTGCTGTATGAAGGCATCAGTCTCAC 1296  
 QY 461 MetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSerAs 480  
 Db 1297 ATGAATCCAAACTCTCTGAATGGTTC-CCAGCGGTGCTGATGAGCGGTAATCTCTGA 1355  
 QY 480 pHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTy 500  
 Db 1356 CCACGAGTGTCCAGAGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGCTA 1415  
 QY 500 rArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLy 520  
 Db 1416 CCGGGTCAGACCTCCCCACATACCTCCCAATGAAGAATTCATGGTGTCACTGCTTCAA 1475  
 QY 520 salalLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAs 540  
 Db 1476 AGCCCTCAAGCTGGGGCAAGAAGGCAAAAGTTCTCTGCAAGAGTGGCCACCTGTACTCAA 1535  
 QY 540 nValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTy 560  
 Db 1536 CGTGACGAGAGAGGTGGCGGCATCATGGAGTCTTACTTCCGCTCGGATACGCCCTCTA 1595  
 QY 560 rPheSerTyrSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArgLy 580  
 Db 1596 CTTTTCTACTCTCATCTGGTGTGCCGCACTGCCATCGAAGAGGTCCAGGCAGAGAGAA 1655  
 QY 580 sAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCy 600

Db 1656 CGATGATAGTCATCCAGTCCAGCTGGACAACTGCATCCTGAATCCGAGACCCTCGTGTG 1715  
Qy 600 sValylsGluProProAlaThrThrPheArgAspTyrSerAlaIleLeuTyrIleuAsnGI 620  
Db 1716 TGTCAAGAGAGCCCGAGCTACACTTCGCGGACTACAGCGCCATCCTTTACCTAAATGG 1775  
Qy 620 yAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGI 640  
Db 1776 GGACTTCGATCGCGGAACCTTTTATTTTACCTGAATCGATGCCAAGACCCTGACGGGAGA 1835  
Qy 640 uValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGI 660  
Db 1836 GGTGAGCCTCAGTGTGGAGAGCGGTGGATCTCTTCAGGCACTGAAACCCACATGG 1895  
Qy 660 yValylsAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrIleuAspPr 680  
Db 1896 AGTGAAGCTGTCCAGAGGGGAGCGCTGTGCCATCGCTGTGGTTCACCTGGACCC 1955  
Qy 680 oArgHisSerGluArgAspArgValGlnAlaAspLeuValylsMetLeuPheSerPr 700  
Db 1956 TCGACACAGCGAGCGGACAGGGTGCAGGACAGTGAATCGAGATGCTCTTCAGCCC 2015  
Qy 700 oGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluPr 720  
Db 2016 AGAAGAGATGGACCTCTCCAGGAGCAGCCCTGGATGCCAGCAGGGTCCCGCCGAACC 2075  
Qy 720 oAlaGlnGluSerLeuSerGlySerGlySerGlySerProLysAspGluLeu 736  
Db 2076 TGCACAGAGTCTCTTCAGGACAGTGAATCGAAGCCCAAGATGAGCTA 2124

## RESULT 14

ABZ11345  
ID ABZ11345 standard; cDNA; 2152 BP.

XX AC ABZ11345;

XX DT 20-JAN-2003 (first entry)

XX DE Human polynucleotide SEQ ID NO 227.

XX KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
cell-proliferative disorder; neurodegenerative disease; bacterial;  
Parkinson's disease; Alzheimer's disease; autoimmune disease;  
multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
arthritis; cyostatic; immunomodulator; neotropic; dermatological;  
antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
antiarthritic; gene; ss.

XX OS Homo sapiens.

XX FN WO200270539-A2.

XX PD 12-SEP-2002..

XX PF 05-MAR-2002; 2002WO-US005095.

XX PR 05-MAR-2001; 2001US-00799451.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao OA, Ren F;  
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
PI Wehrman T, Wang J, Wang D, Drmanac RT;

XX DR WPI; 2002-759812/82.

XX DR P-ESDB; ABP69128.

XX PT New polynucleotides comprising sequences assembled from expressed  
PT sequence tags (ESTs), useful for treating cell-proliferative,  
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet  
or coagulation disorders.

XX

PS Claim 1; SEQ ID NO 227; 1012pp + Sequence Listing; English.  
XX The invention relates to an isolated polynucleotide (I) comprising a  
nucleotide sequence selected from any of 948 sequences (ABZ11119-  
ABZ12066) or their mature protein coding portion, active domain coding  
protein or complementary sequences. The polynucleotides are useful for  
identifying expressed genes or for physical mapping of human genome. The  
encoded polypeptides (ABP6902-ABP6949) are useful as molecular weight  
markers, as a food supplement, for generating antibodies, in medical  
imaging, screening and diagnostic assays and for treating cell-  
proliferative disorders (cancer), neurodegenerative diseases (Parkinson's  
or Alzheimer's disease), autoimmune diseases (multiple sclerosis,  
diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,  
platelet or coagulation disorders, wound, burns, incision, ulcers, liver  
or lung fibrosis, infections (bacterial, viral, fungal, parasitic),  
arthritis, etc. Note: The sequence data for this patent did not form part  
of the printed specification, but was obtained in electronic format  
directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 2152 BP; 494 A; 636 C; 602 G; 420 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.: 4.8e-232 Length: 2152  
Score: 244.00 Matches: 358  
Percent Similarity: 98.90% Conservative: 0  
Best Local Similarity: 98.90% Mismatches: 2  
Query Match: 33.15% Indels: 4  
DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x ABZ11345 (1-2152)

Qy 1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaSer 20  
Db 32 ATGGCGGTACGGCGCTTGAAGTGTGTGACACACTGTGGTGTGTGGCGCTGCCTCC 91  
Qy 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40  
Db 92 CAAGCGAGGTGAGTCCGAGGAGGATGGGGCATGTGTGACGCTGATCTCTCTTGGCC 151  
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
Db 152 GAGGGAGCGGAGCTTACGCGCGGGGACTGGCCGGGTGTCTTCTGAGCATGGAACGG 211  
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
Db 212 GCGCTGCGCTCCCGGAGCGCTCCGCGCCCTTCCCTGCGCTCCGCGCACCCAGTGTGCC 271  
Qy 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProAlaGlnAlaSerGly 100  
Db 272 GCGGACTTCCGTGGAGCTGGACCCCGACTGTGTCCTCCCGCGGCGCCGCGCTCGGGC 331  
Qy 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLe 120  
Db 332 GCGGC-CGCGCTTCCGCGACCTGAGCTTCTTGGGGGCGCTTCTGCGTCTGCGCTGCGCT 390  
Qy 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe 140  
Db 391 GCGCGCTGCTTCCGCT 450  
Qy 140 eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysLeuLeuGly 160  
Db 451 CGCAAGCGGAGCGCTCAACTACCTACCTGCGCTTCTTCAAGATCAACAAGTTGGA 510  
Qy 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGI 180  
Db 511 GAAAGCTGTGTGCGAGCACACCTTCTTCTGGGGCAATCTTGAGGCATCTGGAATGCA 570  
Qy 180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValylsGluAlaAspPheLysAspLe 200  
Db 571 GCAGAACCTAGACTATTACCAAAACCTATCTGGAGTGAAGGAGGCGCGCTTCAAGGATCT 630  
Qy 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGI 220

Db 631 TGAGACTCAACCCATATGCAAGATTTTCGACTGGGAGTGCAGCTCTACTCAGAGGAACA 690  
Qy 220 nProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTyrPheValAlaTyrG1 240  
Db 691 GCCACAGGAAGCTGTGCCCCACCTAGAGCGCGCTGCAGAGATATCTTTGTGGCCCTATGA 750  
Qy 240 uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
Db 751 GGAGTGCCTGCTTCGCAAGGGCCCTATGACTACGATGGCTCAACACTACCTTGAAGTA 810  
Qy 260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysIysG1 280  
Db 811 CAACGCTGACCTTTCAGAGCCATCAGATCATTACATCCAGTCTCAACTGTAAAGCA 870  
Qy 280 nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
Db 871 GAACGTGTACAGGAGCTTCTCCACCCCAAGTCGAGAGAGCCCTTTGAAGACTTCT 930  
Qy 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrAsnIleGlyAsnTyrThrGlnAl 320  
Db 931 CCCATCGCATTAATATATCGAGTTTGCCTACTATAACATTTGGGAATATACACAGGC 990  
Qy 320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
Db 991 TGT-TGAATGTGCAGAGACCTATCTTCTTCTTCCCAATGACAGGATGATGAACCAA 1049  
Qy 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArg 360  
Db 1050 ATTATGGCTATTATGACGCTATGCTGGAGAGAACACACACAGATCATCGGCCCGCTG 1109  
Qy 360 lu 360  
Db 1110 AG 1111

RESULT 15  
ABZ11344  
ID ABZ11344 standard; cDNA; 1309 BP.  
XX AC ABZ11344;  
XX DT 20-JAN-2003 (first entry)  
XX DE Human polynucleotide SEQ ID NO 226.  
XX KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;  
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;  
KW antiarthritis; Gene; ss.  
XX OS Homo sapiens.  
XX WO200270539-A2.  
XX PN 12-SEP-2002.  
XX PD 05-MAR-2002; 2002WO-US0005095.  
XX PR 05-MAR-2001; 2001US-00799451.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AU, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
PI Wehrman T, Wang J, Wang D, Drmanac RT;  
XX WPI; 2002-759812/82.  
XX P-PSDB; ABP69127.  
XX New polynucleotides comprising sequences assembled from expressed

PT sequence tags (ESTs), useful for treating cell-proliferative,  
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet  
XX or coagulation disorders.  
PS Claim 1; SEQ ID NO 226; 1012pp + Sequence Listing; English.  
XX The invention relates to an isolated polynucleotide (1) comprising a  
CC nucleotide sequence selected from any of 948 sequences (ABZ11119-  
CC ABZ12066) or their mature protein coding portion, active domain coding  
CC protein or complementary sequences. The polynucleotides are useful for  
CC identifying expressed genes or for physical mapping of human genome. The  
CC encoded polypeptides (ABP6902-ABP6984) are useful as molecular weight  
CC markers, as a food supplement, for generating antibodies, in medical  
CC imaging, screening and diagnostic assays and for treating cell-  
CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's  
CC or Alzheimer's disease), autoimmune diseases (multiple sclerosis,  
CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,  
CC platelet or coagulation disorders, myeloid or lymphoid disorders, liver  
CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),  
CC arthritis, etc. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX Sequence 1309 BP; 287 A; 384 C; 365 G; 273 T; 0 U; 0 Other;  
SQ  
Alignment Scores:  
Pred. No.: 2.12e-207 Length: 1309  
Score: 219.00 Matches: 406  
Percent Similarity: 99.02% Conservative: 0  
Best Local Similarity: 99.02% Mismatches: 2  
Query Match: 29.76% Indels: 4  
DB: 6 Gaps: 0  
US-10-045-815-4 (1-736) x ABZ11344 (1-1309)  
Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
Db 32 ATGGCGGTACGGCGCTTGAAGCTGCTGACACACTGCTGGCTGCTGCGCTCCCTCC 91  
Qy 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40  
Db 92 CAAGCCGAGGTGAGTCCGAGCGAGGATGGGCGCATGGTGACGCTGATCTCTCTCC 151  
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
Db 152 GAGGGGACCGGACCTACGCGCGGGGACTGGCCCGGGTGGTCTCTGAGCATGGACGG 211  
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
Db 212 GCGCTGCGCTCCGCGGACGCTCCGCGCCCTTCCGCTGCGCTGCCGACCCAGTGTCC 271  
Qy 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
Db 272 GCCGACTTCCGCTGGGAGCTGACCCCGACCTGGTCCCGCCGCGCCAGGCTTCGGGC 331  
Qy 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLe 120  
Db 332 GCGGC- GCGCTGCGCGACCTGAGCTTCTTCGGGGGCTCTTCTGCTGCGCTGCTGCT 390  
Qy 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140  
Db 391 GCGCGCTGCTCCGCGCGCGCGCGCCCTACCTGAGGTGCGCTTCTCAAGATCAACAGT 450  
Qy 140 eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuG1 160  
Db 451 CGCGAAGCGGAGCCCTACACTACCTGAGGTGCGCTTCTCAAGATCAACAGTTCGA 510  
Qy 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetG1 180  
Db 511 GAAAGCTGTGCTGCAGCACACACCTTCTTCTGCGGCAATCTCTGAGCACATGGAAATGCA 570  
Qy 180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200

Db 571 GCAGAACTAGACTATTACCAACCATGCTCTGGAGTGAAGAGGCCGACTTCAAGGATCT 630  
Qy 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGlu 220  
Db 631 TGAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTGGACTCTACTCAGAGGAACA 690  
Qy 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGl 240  
Db 691 GCCACAGGAAGCTGTGCCACCTAGAGGGGGCTGCAAGAATACCTTTGTGGCCCTATGA 750  
Qy 240 uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
Db 751 GGAGTGGCGTGCCTCTGGAGAGGGCCCTATGACTACGATGGCTACAACTACCTTGAGTA 810  
Qy 260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGl 280  
Db 811 CAAGCTGACCTCTTCCAGGCCATCACAGATCATTACATCCAGTCCCTCAACTGTAAGCA 870  
Qy 280 nAsnCysValThrGlnLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
Db 871 GAAGTGTACGGAGCTTGCTTCCCAAGTCGAGAGAGCCCTTTGAAGACTTCCT 930  
Qy 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320  
Db 931 CCCATCGCATTATTAATTATCTGCAGTTTGCTACTATAACATTGGGAATTTATACAGGC 990  
Qy 320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
Db 991 TGT-TGAATGCGCAAGACCTATCTTCTCTTCCCAATGACGAGGTGATGAACCAA 1049  
Qy 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG 360  
Db 1050 ATTTGGCCTATTATGCAGCTATGCTTGAGAGAAGACACACAGATCCATCGGCCCCGCTG 1109  
Qy 360 luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaT 380  
Db 1110 AGAGTGCAAGGAGTACCGACAGCGAGCCCTACTGGAAAAAGAACTGCTTTCTTCGCTT 1169  
Qy 380 yrAspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIleP 400  
Db 1170 ATGATGTTTTGGAAATTCCTTTGTGGATCGGATTCATGGACTCCAGAGAAAGTGATTC 1229  
Qy 400 roLysArgLeuGlnGluLysGlnLys 408  
Db 1230 CCAAGAGATTGCAAGAGAAACAGAAG 1255

Search completed: July 18, 2004, 18:38:50  
Job time : 1051 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 18:22:31 ; Search time 1081 Seconds  
(without alignments)  
3321.616 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 736  
Sequence: 1 MAVRALKLLTLLAVVAAS.....PPEPAQESLSGSEKPKDEL 736

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3190992 seqs, 2439311697 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6373542

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool\_P/US10045815/runat\_14072004\_123120\_17294/app\_query.fasta\_1.903  
-DB=Published Applications NA -QFWT=fastap -SURFIX=oli.rnpb -MINMATCH=0.1  
-LOOFCU=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1  
-ALIGN=15 -MODE=LOCAL -OUTFWT=pto -NORM=ext -HRAPISE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10045815 @CCN 1 1 912 @runat 14072004\_123120\_17294  
-NCPU=6 -ICFU=3 -NO.MAP -LARGECURY -NEG SCORES=0 -WAIT -DSPLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Published Applications NA: \*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq: \*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq: \*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq: \*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq: \*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq: \*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq: \*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq: \*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq: \*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq: \*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq: \*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq: \*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq: \*  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq2: \*  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq: \*  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq: \*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq: \*  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq: \*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq: \*  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
------------	-------	--------------	----	-------------

ALIGNMENTS

RESULT 1

US-10-045-815-3  
; Sequence 3, Application US/10045815  
; Publication No. US20020160498A1  
; GENERAL INFORMATION:  
; APPLICANT: Madhwa, Renu  
; APPLICANT: Sugihara, Takashi  
; APPLICANT: Chide, Akiko  
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
; FILE REFERENCE: 06501-091001  
; CURRENT APPLICATION NUMBER: US/10/045,815  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: PCT/JP00/02731  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: JP 11/118806  
; PRIOR FILING DATE: 1999-04-26  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 2600  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:

Sequence 3, Appli  
Sequence 228, App  
Sequence 2, Appli  
Sequence 69, Appl  
Sequence 28, Appl  
Sequence 51, Appl  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 227, App  
Sequence 226, App  
Sequence 6419, Ap  
Sequence 6419, Ap  
Sequence 7, Appli  
Sequence 5, Appli  
Sequence 2327, A  
Sequence 574, App  
Sequence 708, App  
Sequence 708, App  
Sequence 8286, Ap  
Sequence 15266, A  
Sequence 2337, Ap  
Sequence 2337, Ap  
Sequence 3847, Ap  
Sequence 11074, A  
Sequence 13025, A  
Sequence 1401, Ap  
Sequence 123, App  
Sequence 508, App  
Sequence 1793, Ap  
Sequence 217, App  
Sequence 86, Appl  
Sequence 921, App  
Sequence 4475, Ap  
Sequence 618, App  
Sequence 279034,  
Sequence 279035,  
Sequence 279036,  
Sequence 279037,  
Sequence 279034,  
Sequence 279035,  
Sequence 279036,  
Sequence 279037,  
Sequence 8942, Ap  
Sequence 158857,  
Sequence 158859,

NAME/KEY: CDS  
LOCATION: (52)... (2259)  
US-10-045-815-3

## Alignment Scores:

Pred. No.: 0 Length: 2600  
Score: 736.00 Matches: 736  
Percent Similarity: 100.00% Conservativeness: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 14 Gaps: 0

US-10-045-815-4 (1-736) x US-10-045-815-3 (1-2600)

```
QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20
DB 52 ATGGCGGTACCGCGGTTGAAGCTGCTGACCACTGCTGGCTGCTGGCGGCTGCCCTC 111
QY 21 GlnAlaGluValGluSerGluAlaGlyTyrProGlyMetValThrProAspLeuLeuPheAla 40
DB 112 CAAGCCGAGGTCGAGTCGAGCGGAGGATGGGCGATGGTGCAGCGCTGCTCTCGCC 171
QY 41 GlnGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60
DB 172 GAGGGGACCGGAGCTACCGCGCGGGGACTGGCGCGGGTGGTCTCTGAGCATGGAGACGG 231
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
DB 232 CGCTGCGCTCCGGGAGCCCTCGCGCCCTCGCTGCGCTGCGCTGCGCCGACCCAGTGTGCC 291
QY 81 AlaAspPheProTyrGluLeuAspProAspTyrProSerProAlaGlnAlaSerGly 100
DB 292 GCCGACTTCCGCGGAGCTGGACCCCGACTGCTGCCCGCCGCGCGCCGAGGCGCTGGGC 351
QY 101 AlaGlyAlaLeuArgAspLeuSerPheGlyClyLeuLeuArgArgAlaAlaCysLeu 120
DB 352 CGCGCGCCCTCGCGGACCTGAGCTTCTTCGGGGGCGCTTCTGCGTCCGCTGCGCTGCGCTG 411
QY 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluLeuMetGluLeuGluPhe 140
DB 412 CGCGCTGCTCGCGCGCGCGCGCGCCACTCGCTCGCTCAGCGAGAGATGGAGTGGAGTTC 471
QY 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160
DB 472 CGCAAGCGGAGCCCTACAACTACTCGAGGTGCGCTACTCAAGATCAACAGTGGAG 531
QY 161 LysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetGln 180
DB 532 AAAGCTGTGTGTCAGCACACACTTCTTCGTGGGCAATCTTGAGCACATGGAAATGCAG 591
QY 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200
DB 592 CAGAACCTAGACTATTACCAACCATCTCTGGAGTGAAGGAGCGCGACTTCAAGGATCTT 651
QY 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220
DB 652 GAGACTCAACCCCATATGCAAGANTTTGACTGGAGTGGAGTCTACTCTCAGAGGAGACAG 711
QY 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240
DB 712 CCACAGGAAGTGTGGCCCACTAGAGGCGCGCTGCAAGATACTTTTGGGCTATGAG 771
QY 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260
DB 772 GAGTGGCGTCCCTCTCGGAAGGCCCTATGACTACGATGCTACACTACCTTGTAGTAC 831
QY 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280
DB 832 AACGCTGACCTCTTCCAGGCCATCACAGATCATATACATCCAGGTCCTCAACTGTAAAGCAG 891
QY 281 AsnCysValThrGluLeuAlaSerHisProSerArgLysProPheGluAspPheLeu 300
DB 892 AACTGTGTACGGAGCTTGTCTTCCACCCCAAGTCGAGAGGCCCTTTGAAGACTTCCCTC 951
```

```
QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320
DB 952 CCATCGCATTAATAATTATCTGCAGTTTGCCCTACTATAACATTTGGGAATTATACCAAGCT 1011
QY 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340
DB 1012 GGTGAATGTGCCAGACCTTCTTCTTCCCAATGACGAGGTGATGAACCAAT 1071
QY 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360
DB 1072 TTGGGCTATTATGCAGCTATGCTTGGAGAACAACACCATCATCCATCGGCCCGCTGAG 1131
QY 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr 380
DB 1132 AGTGCCAGGAGTACCGACAGCAAGCCTCTGGAAGAAGAACTGCTTTTCTTGCCTTAT 1191
QY 381 AspValPheGlyIleProPheValAspProAspSerTyrProGluGluValIlePro 400
DB 1192 GATGTTTTTGGAAATCCCTTTTGGATCCGGATTCTGCAGCTCCAGAAAGAGTGTATCCC 1251
QY 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420
DB 1252 AAGAGATTGCAAGAGAAACAGCAAGTCAAGACGGGAAACAGCGCTACGCATCTCCAGGAG 1311
QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440
DB 1312 ATTGGGAACCTTATGAGGAATCGAGACCTTGTGGAAGAGAGACCAAGAGTCACTG 1371
QY 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460
DB 1372 GATGTGACGAGACTACCCGGGAAAGTGGCGCCCTGCTGTATGAAGGCATCAGTCTCACC 1431
QY 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480
DB 1432 ATGAACTCCAACTCTCTGAATGGTTACCGCGGGTGGTGTGATGGACGGGCTAATCTCTGAC 1491
QY 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500
DB 1492 CACGAGTGTCCGAGAGCTCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGGTAC 1551
QY 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520
DB 1552 CGGGTTCAGACTCCCCACATCTCCCAATGAAGAATCTCTATGTGTGTCACGTCTTCAA 1611
QY 521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540
DB 1612 GCCCTCAAGCTGGGGCAAGAGGCAAAAGTTCTCTGCAGAGTGGCCACCTGTACTACAAC 1671
QY 541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560
DB 1672 GTGCGGAGAAAGTGGCGGCTCATGAGTCTTCTTCCGCTGGATGACCCCTCTAC 1731
QY 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580
DB 1732 TTTTCTCTCTCTCATCTGTGTGCGCACTGCCATCGAAGAGGTCCAGGACAGAGGAAG 1791
QY 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600
DB 1792 GATGATAGTCTCCAGTCCAGTCCAGCAACTGCACTCTGAAATGCCGAGACCTCGTGTGT 1851
QY 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620
DB 1852 GTCAAGAGAGCCCCAGCCCTACACCTTCGCGACTACAGCGCATCTTACTTAATGGG 1911
QY 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640
DB 1912 GACTTCGATGGCGAAACTTTTATTCTACTGAATGCAAGTGCACAGCCGTGACGCGAGAG 1971
QY 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660
DB 1972 GTGCAAGCTCTAGTGTGGAGAGCGGTGGATTTCTTTCAGGCATCTGAAACCCACATGGA 2031
```

QY 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680  
DB 2032 GTAAAGGCTGTACACAGGGGAGCGCTGTGCCATCGCCTGTGGTTACCTTGACCCCT 2091  
QY 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
DB 2092 CGACACAGCGAGCGGACAGAGGTTGAGGAGATGACCTGGTGAAGATGCTCTTCAGCCCA 2151  
QY 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProGluPro 720  
DB 2152 GAAGAGATGGACCTCTCCAGGAGCAGCCCTGGATGCCAGAGGGCCCCCGGAACCT 2211  
QY 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
DB 2212 GCACAGAGTCTCTCTCAGGCACTGAATCGAAGCCCCAAGGATGAGCTA 2259

## RESULT 2

US-10-302-172-228  
; Sequence 228, Application US/10302172  
; Publication No. US20040053250A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. US20040053250A1 Arginine-rich Protein-like Nucleic Acids an  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 803.1CNCIP  
; CURRENT APPLICATION NUMBER: US/10/302,172  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 10/235,251  
; PRIOR FILING DATE: 2002-08-20  
; PRIOR APPLICATION NUMBER: PCT US02/05095  
; PRIOR FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: US 09/799,451  
; PRIOR FILING DATE: 2001-03-05  
; NUMBER OF SEQ ID NOS: 950  
; SOFTWARE: pt\_FL\_genes Version 2.0  
; SEQ ID NO 228  
; LENGTH: 2242  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (32)..(2239)  
US-10-302-172-228

Alignment Scores:  
Pred. No.: 0 Length: 2242  
Score: 634.00 Matches: 734  
Percent Similarity: 99.46% Conservative: 0  
Best Local Similarity: 99.46% Mismatches: 2  
Query Match: 86.14% Indels: 4  
DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x US-10-302-172-228 (1-2242)

QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
DB 32 ATGGCGGTACGCGGTTGAAGCTGCTGACCACACTGTGGCTGTGCGGCGCTGCTCC 91  
QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40  
DB 92 CAAGCCGAGGTGAGTTCGAGGAGAGATGGGGCATGGTGACGCTGATCTCTTCGCC 151  
QY 41 GluGlyThrAlaAlaTyAla-ArgGlyAspTrpProGlyValValLeuSerMetGluAr 60  
DB 152 GAGGCGACCGCAGCCTACTC-GCGCGGGAGCTGGCCCGGGGTGGTCTCGATGGAACG 210  
QY 60 gAlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAl 80  
DB 211 GGCCTCGCTCCCGGGAGCCCTCCCGCCCTTCGCTGGCGCCGCCACCCAGTGTC 270  
QY 80 aAlaAspPheProTrpGluLeuAspProAspTrpSerProAlaGlnAlaSerGl 100

DB 271 CGCGACTTCCGTGGAGCTGGACCCGACTGTGTCCCGACCGCGCCCGAGGCTCGG 330  
QY 100 yAlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuArgArgAlaAlaCysL 120  
DB 331 CGCGCGC-CGCGCTCGCGACCTGAGCTTCTTGGGGGCCCTTCTGCTGCGCTCGCTGCC 389  
QY 120 euArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluP 140  
DB 390 TGGCGCGCTGTCTGGCGCGCGCGCGCCCACTCGCTCAGCGAAGAGATGGAGTGGAGT 449  
QY 140 heArgLysArgSerProTrpAsnTyLeuGlnValAlaTyPheLysLysLeuLysLeuG 160  
DB 450 TCCGCAAGCGGAGCCCTTACAACTACTCTGAGGTGCGCTACTTCAAGATCAACAAGTTGG 509  
QY 160 luLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetG 180  
DB 510 AGAAAGCTGTGTCTGTCAGCACACACCTTCTTCTGGGGCNAITCCTGAGCACATGGNAATGC 569  
QY 180 lnGlnAsnLeuAspTyTrpGlnThrMetSerGlyValLysGluAlaAspPheLysAspL 200  
DB 570 AGCAGAACCTAGACTATTACCAACCATGTCTGGAGTGAAGAGCGCCGACTTCAAGGATC 629  
QY 200 euGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyTrpSerGluGluG 220  
DB 630 TTGAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTGGACTCTTACTCAGAGGAAC 689  
QY 220 lnProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyTrpPheValAlaTyR 240  
DB 690 AGCCACAGAGAGCTGTGCCACCTAGAGGGCGGCTGCAAGAATACITTTGTGGCCTATG 749  
QY 240 luGluCysArgAlaLeuCysGluGlyProTyAspTyAspGlyTyTrpAsnTyTrpLeuGluT 260  
DB 750 AGGAGTGGCGTGGCTCTCGAAAGGGCCCTATGACTACGATGGGTACAACTACCTTGAGT 809  
QY 260 yTrpAlaAlaAspLeuPheGlnAlaIleThrAspHisTyTrpIleGlnValLeuAsnCysLysG 280  
DB 810 ACACGCTGACCTTTCAGGCCATCACAGATCATATCATCCAGTCTCTCACTGTAAAGC 869  
QY 280 lnAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheL 300  
DB 870 AGAAGCTGTGTACGGAGCTTGTCTCCACCCCAAGTCGAGAGAAAGCCCTTGAAGACTTCC 929  
QY 300 euProSerHisTyTrpAsnTyLeuGlnPheAlaTyTrpAsnIleGlyAsnTyTrpGlnA 320  
DB 930 TCCCATCGCATTAATAATTATCTGCAGTTGCTTACTATAACATTGGGAATTATACACAG 989  
QY 320 laGlyGluCysAlaLysThrTyTrpLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
DB 990 CTGGTGATGTGCCAAGACCTATCTTCTTCTTCCCAATGACGAGGTGATGAACCAA 1049  
QY 340 snLeuAlaTyTrpAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG 360  
DB 1050 ATTTGGCCTATTATGCTAGCTATGCTTGGAGAAGAACACACCAGATCCATCGGCCCCCTG 1109  
QY 360 luSerAlaLysGluTyTrpArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaAr 380  
DB 1110 AGAGTGCCAAGAGTACCGACAGCGAGCCTACTCTGAAAAAGAACTGTTTTCTTCGCTT 1169  
QY 380 yTrpValPheGlyIlePhePheValAspProAspSerTrpThrProGluGluValIleP 400  
DB 1170 ATGATGTTTTTGAATTTCTTGTGGATCCGGATTCATGGACTCCAGAAAGTGAITTC 1229  
QY 400 roLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
DB 1230 CCAAGAGATTGCAAGAGAAACAGAAAGTCAGAACGGGAAACACCGCTACCATCTCCAGG 1289  
QY 420 luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440  
DB 1290 AGATTGGGAACCTTATGAAGGAAATCGAGACCCCTTGTGGAGAGAAGAACCAAGGAGTGC 1349  
QY 440 euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyTrpGluGlyIleSerLeuT 460

```
Db 1350 TCGATGTGAGCAGACTGACCGGAGAGTGGCCCTGCTGTATGAGGCACTAGTCTCA 1409
Qy hrMetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerA 480
Db 1410 CCATGAACCTCAAACTCTGAATGTTACCAAGCGGTGGTGATGACGGCGTAACTCTG 1469
Qy spHisGluCysGlnLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyT 500
Db 1470 ACCAGAGTGTGAGAGCTGCAGAGACTGACCAATGTGGCAGCAACTTCAGGAGATGGCT 1529
Qy yrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheL 520
Db 1530 ACCGGGTTCAGACCTCCCACTACTCCCAATGAAGATTCATATGGTGTCACTGTCTTCA 1589
Qy ysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrA 540
Db 1590 AAGCCCTCAAGCTGGGCAAGAGGCAAAAGTTCCTCTGCAGAGTGGCCACCTGTACTACA 1649
Qy snValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuT 560
Db 1650 ACGTGACGGAGAAAGTGGCGGCATCATGGAGTCTACTTCGGCTCGATACGCCCTCT 1709
Qy yrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValAlaGluArgL 580
Db 1710 ACTTTTCTACTCTCATCTGTGTGGCGCACTGCATCGAAGAGGTCCAGGCAAGAGGA 1769
Qy ysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValC 600
Db 1770 AGGATGATAGTATCCAGTCCAGTGCAGCAACTGCATCTCGAATGCCGACCCCTGTGT 1829
Qy ysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnG 620
Db 1830 GTGTCAAGAGAGCCCCAGCTTACACCTTCGCGACTACAGCGCCATCTTTACCTAAATG 1889
Qy lyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaG 640
Db 1890 GGGACTTCGATGGCGGAACTTTATTTCATCTGAATCGGATGCCAGACCGTGACGCGAG 1949
Qy luValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisG 660
Db 1950 AGGTGACGCTCAGTGTGGAAGAGCCGTGGGATTCCTTCAGGCACTGAAACCCACATG 2009
Qy lyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuAspP 680
Db 2010 GAGTGAAGGTGTACCAAGGGGAGCGCTGTGCAATCGCCCTGTGCTTACCCCTGAGCC 2069
Qy roAtsHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerP 700
Db 2070 CTCGACACAGCGAGCGGAGCAGGCTGAGGAGATGACCTGGTGAAGATGCTCTTCAGCC 2129
Qy roGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGluP 720
Db 2130 CAGAAGAGATGAGACTCTCCAGAGAGCAGCCCTGTGATGCCAGAGGCGCCCGCGAAC 2189
Qy roAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736
Db 2190 CTGCAAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2239
```

## RESULT 3

```
US-10-257-174-2
; Sequence 2, Application US/10257174
; Publication No. US20040034194A1
; GENERAL INFORMATION:
; APPLICANT: Agarwal, Pankaj
; APPLICANT: Murdoch, Paul R.
; APPLICANT: Rizvi, Safia K.
; APPLICANT: Smith, Randall F.
; APPLICANT: Xiang, Zhaoying
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50022
; CURRENT FILING DATE: 2002-10-10
; PRIOR APPLICATION NUMBER: US/10/257,174
; PRIOR APPLICATION NUMBER: PCT/US01/11797
```

```
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/196,603
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/199,417
; PRIOR FILING DATE: 2000-04-24
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2211
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-257-174-2
```

## Alignment Scores:

```
Pred. No.: 0 Length: 2211
Score: 433.00 Matches: 733
Percent Similarity: 99.19% Conservative: 0
Best Local Similarity: 99.19% Mismatches: 3
Query Match: 58.83% Indels: 6
DB: 13 Gaps: 0
```

US-10-045-815-4 (1-736) x US-10-257-174-2 (1-2211)

```
Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaAlaSer 20
Db 1 ATGGCGGTACCGCGCTTGAAGCTGCTGACACACACTGCTGGCTGCTGGCGCTGCCCTCC 60
Qy 21 GlnAlaGluValGluSerGluAlaGlyTropGlyMetValThrProAspLeuPheAla 40
Db 61 CAAGCCGAGGTGAGTCCGAGGAGGATGGGCGATGGTACGCTGATCTCTCTCTGCC 120
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60
Db 121 GAGGGAGCGGAGCCCTACGCGCGGGGACCTGGCCCGGGTGGTCTTGAGCATGGACGG 180
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
Db 181 CGCGTGGCTCCCGGCGAGCCCTCCGCGCCCTTCGCGCTGCGCTGCGCCAGCTGTGCC 240
Qy 81 AlaAspPheProTyrGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100
Db 241 GCGGACTTCCCGTGGAGCTGGAACCCGACTGTGTCCTCCAGCCGCGCCGCTCGGCG 300
Qy 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLe 120
Db 301 GCGCGC-CGCGCTCGCGACCTGAGCTTCTTCGGGGGCTTCGCGCTGCGCTGCGCTGCT 359
Qy 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluMetGluLeuGluPh 140
Db 360 GCGCGCTGCTCGGGCGCGCGCCCTGCTGAGTGGAGTGGAGGAGGAGGAGGAGTTCG 419
Qy 140 eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGl 160
Db 420 CGCGAAGCGGAGCCCTACAACTACCTGAGCTGCGCTTCAAGATCAACAGTTGGA 479
Qy 160 uLysAlaValAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGl 180
Db 480 GAAAGCTTCTGCTGAGCAGACACACTTCTTCGTTGGGCAATCTCTGAGCAGATGGA 539
Qy 180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200
Db 540 GCAGAACCTAGACTATTACCAACCATGTCTGGAGTGAAGAGGAGGAGGAGGAGTCT 599
Qy 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGl 220
Db 600 TGAGACTCAACCCCATATGCAAGAAATTTTCGACTGGGAGTGGCACTCTACTCAGAGGA 659
Qy 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGlnTyrPheValAlaTyrGl 240
Db 660 GCCAGAGAGTGTGCCCCACCTAGAGGGCGGCTGCAAGATATCTTTGGGCTTATGA 719
Qy 240 uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260
```

720 GGAGTGGCGTCCCTCTGCGAAGGGCCCTATGACTACGATGGCTACAACACTACCTTGAGTA 779  
Qy rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysG1 280  
Db CAACGCTGACCTCTTCAGGSCCATCACAGATCATTTACATCCAGTCTCTCAACTGAAGCA 839  
Qy nAsnCysValThrGluLeuAlaSerHisProSerArgGluGluLysProPheGluAspPheLe 300  
Db GAACTGTGTACAGAGCTTGTCTTCCACCCCAAGTCGAGAGAGCCCTTTGAAGACTTCT 899  
Qy uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320  
Db CCCATCGCATTTATATATCTGAGTTTGCCTACTATATACATTTGGGAATATACACAGGC 959  
Qy agly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
Db TGT-TGAATGTGCGAAGACTATCTCTCTTCTTCCCAATGACGAGGTGATGAACCAA 1018  
Qy snLeuAlaTyrTyrAlaAlaMetLeuGlyGluHisThrArgSerIleGlyProArgG 360  
Db ATTGGCCTATTATGAGCTATGCTTGGAGAAGAACACACAGATCCATCGGCCCCCGTG 1078  
Qy lueSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAla 380  
Db AGAGTGCACAGGAGTACCGACACGAGCCCTACTGGAAGAAAGAACTGCTTTTCTCGCTT 1138  
Qy YrAspValPheGlyIleProPheValAspProAspSerThrThrProGluGluValIleP 400  
Db ATGATGTCTTTTGAATTTCCCTTTGTGATCCGATTCATGACTCCAGAGAGTGAATTC 1198  
Qy rLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgLysSerGlnG 420  
Db CCAAGAGATTGCAAGAGAACACAGATCAGACGGGAAACAGCCGTACCGATCTCCAGG 1258  
Qy lueGlyAsnLeuMetLysGluLeuThrLeuValGluGluLysThrLysGluSerL 440  
Db AGATTGGAACTTATGAGAAATATGAGAAATGAGACCTTGTGGAAGAGAAGAACCAAGAGTCA 1318  
Qy euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeu 460  
Db TGGATGTGACGAGACTGACCCGGGAGGTGGCCCTCTGTGTATGAGGATCAGTCTCA 1378  
Qy hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer 479  
Db CCATGAACTCCAACTCTCTGAATGTTTC-CCAGCGGTGTGTGATGAGCGCGTAATCTCT 1437  
Qy AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499  
Db GACCACAGGTGTACGAGAGCTGCAGAGACTGACCAATGTGCGACCAACCTCAGAGATGGC 1497  
Qy TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe 519  
Db TACCGGGGTGAGACTCCACATACATCTCCCAATGAAAGTTCTATGGTGTCACTGTCTTC 1557  
Qy LysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyr 539  
Db AAAGCCCTCAAGCTGGGCGAAGGCAAGGAAAGTTCTCTGAGAGTGCACACCTGTACTAC 1617  
Qy AsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeu 559  
Db AACGTGACGAGAGGTCGCGCGATCATGAGTCTTCTCCCTGGATAGCCCTC 1677  
Qy TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArg 579  
Db TACTTTTCTCTCTATCTGTTGTCGCGACTGCCATCGAAGAGTTCAGGCGAGAGG 1737  
Qy LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599  
Db AAGGATGATAGTATCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGT 1797  
Qy CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619  
Db TGTGTCAAAGAGGCCCCAGGCTACACCTTCGCGACTACAGCGCCATCTTTACCTAAAT 1857

Qy GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla 639  
Db GGGGACTTCGATGGCGGAAACTTTTATTTCACTGAATGGATGCAAGACCGTGCAGCA 1917  
Qy GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659  
Db GAGTGTGAGCTCTAGTGTGGAAGAGCCGTGGATTTCTTTCAGGACCTGAAAACCCACAT 1977  
Qy GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAsp 679  
Db GGAGTGAAGGTGTCTCACCGGGGCGAGCTGTGCCATCGCCCTGTGGTTTCACTCCCTGAC 2037  
Qy ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer 699  
Db CTTGACACACGAGCGGAGCAGGTCGAGGATGATGATGATGATGATGATGATGATGATGAT 2097  
Qy ProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGlu 719  
Db CCAGAGAGATGGACTCTCCAGGAGCAGCCCTGGATGCCAGCAGGTCCTTTCAGC 2157  
Qy ProAlaGlnGluSerLeuSerGlySerGlySerGlySerGlySerGlySerGlySerGly 736  
Db CCTGCACAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2208

## RESULT 4

US-10-312-352-69  
; Sequence 69, Application US/10312352  
; Publication No. US20040053824A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.; TANG, Y. Tom  
; APPLICANT: YUE, Henry; AZIMZAI, Yalda  
; APPLICANT: HE, Ann; BATRA, Sajeev  
; APPLICANT: LO, Terence P.; NGUYEN, Damien B.  
; APPLICANT: BURRILL, John D.; MARCUS, Gregory A.  
; APPLICANT: ZINGLER, Kurt A.; GANDHI, Ameena R.  
; APPLICANT: LAL, Preeti G.; KEARNEY, Liam  
; APPLICANT: BURFORD, Neil; YAO, Monique G.  
; APPLICANT: CHAWLA, Narinder K.; ELLIOT, Vicki S.  
; APPLICANT: ARVIZU, Chandra S.; KHAN, Farrah A.  
; APPLICANT: BAUGHN, Mariyah R.; HAPALIA, April, J.A.  
; APPLICANT: POLICKY, Jennifer L.; AU-YOUNG, Janice K.  
; APPLICANT: LU, Yan; BOROWSKY, Mark L.  
; APPLICANT: LU, Dying Aina M.; RAMKUMAR, Jayalaxmi  
; APPLICANT: YANG, Junming; GURURAJAN, Rajagopal  
; APPLICANT: WARREN, Bridget A.; GIETZEN, Kimberly J.  
; APPLICANT: XU, Yuming; KALLICK, Deborah A. Kavitha  
; APPLICANT: LEE, Ernestine A.; THANGAVELU, Kavitha  
; APPLICANT: DEEGEANE, Angelo M.; LEE, Sally  
; TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES  
; FILE REFERENCE: PF-0794 USN  
; CURRENT APPLICATION NUMBER: US/10/312,352  
; PRIORITY FILING DATE: 2002-12-18  
; PRIOR APPLICATION NUMBER: PCT/US01/21067  
; PRIOR FILING DATE: 2001-08-29  
; PRIOR APPLICATION NUMBER: US 60/215,454  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 60/219,462  
; PRIOR FILING DATE: 2000-07-18  
; PRIOR APPLICATION NUMBER: US 60/240,111  
; PRIOR FILING DATE: 2000-10-12  
; PRIOR APPLICATION NUMBER: US 60/240,106  
; PRIOR FILING DATE: 2000-10-12  
; PRIOR APPLICATION NUMBER: US 60/244,021  
; PRIOR FILING DATE: 2000-10-27  
; PRIOR APPLICATION NUMBER: US 60/248,887  
; PRIOR FILING DATE: 2000-11-14  
; PRIOR APPLICATION NUMBER: US 60/249,570  
; PRIOR FILING DATE: 2000-11-16  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: PERL Program  
; SEQ ID NO 69  
; LENGTH: 2583

; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: misc feature			
; OTHER INFORMATION: Incyte ID No. US20040053824A1 6780147CB1			
US-10-312-352-69			
Alignment Scores:			
Pred. No.:	0	Length:	2583
Score:	433.00	Matches:	733
Percent Similarity:	99.19%	Conservative:	0
Best Local Similarity:	99.19%	Mismatches:	3
Query Match:	58.83%	Indels:	6
DB:	13	Gaps:	0
US-10-045-815-4 (1-736) x US-10-312-352-69 (1-2583)			
QY	1	MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaAlaSer	20
DB	42	ATGGCGGTACGCGGTGAGCTGCTGACCACTGCTGGCTGTGTGGCGGCTGCTCC	101
QY	21	GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla	40
DB	102	CAAGCCGAGGTCGAGTCCGAGGAGGATGGGCGATGGTGACGCCCTGATCTGCTTCGCC	161
QY	41	GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg	60
DB	162	GAGGGGACCGAGCTACGGCGCGGGGAGCTGGCCCGGGGTGGTCTCTGAGCATGGAACGG	221
QY	61	AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla	80
DB	222	CGCGTGGCTCCCGGGGAGCCCTCCGCGCCCTTCGCTGCTGGCTGCCGACCCAGTGTGCC	281
QY	81	AlaAspPheProTyrGluLeuAspProAspTyrSerProSerProAlaGlnAlaSerGly	100
DB	282	GCCGACTTCGCGTGGGAGCTGGACCCGAGCTGGTTCGCCCGCGCGGCCAGGCTCGGC	341
QY	101	AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLe	120
DB	342	CGCGC-CGCGCTGCGGACCTGAGCTTCTTCGGGGGCTTCGCGTGGCTGCGCTGCCCT	400
QY	120	uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh	140
DB	401	CGCGCGCTGCTCG	460
QY	140	eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuG	160
DB	461	CCGCAAGCGGAGCCCTCAACTACCTGCGAGGTGCGCTTCAAGATCAACAAAGTTGA	520
QY	160	uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetG	180
DB	521	GAAGCTGTGTCGACGACACACCTTCTTCGTGGGCAATCCTGAGCACATGGAATGCA	580
QY	180	nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe	200
DB	581	GCAGAACCTAGACTATTACCAACCACTGCTGGAGTGAAGGAGCGCGACTTCAAGGATCT	640
QY	200	uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluG	220
DB	641	TGAGACTCAACCCCATATGCAAGAATTTGCACTGGGAGTGGCTACTCTACTCAGAGGA	700
QY	220	nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrG	240
DB	701	GCCACAGGAAGCTGTGCCCCACCTAGAGGCGGCGTGCAGAAATACTTTTGGCGCTATGA	760
QY	240	uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluT	260
DB	761	GGAGTCGCGTGCCTCTGCGAAGGCGCCCTATGACTACGATGGCTACAACTTACCTTGA	820
QY	260	rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysG	280
DB	821	CAACGCTGACCTCTCCAGGCGCATCACAGATCATTTACATCCAGGTCTCTCACTGTAAG	880
QY	280	nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe	300
DB	881	GAACGTGTGTACGGAGCTTGTCCACCCCAAGTCGAGAGAAGCCCTTTGAAGACTTCT	940
QY	300	uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl	320
DB	941	CCCATCGCATTAATAATATCTGCAGTTTGGCTTACTATAACATTGGGAATTATACACAG	1000
QY	320	agly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA	340
DB	1001	TGT-TGAATGTCCCAAGACCTATCTTCTTCTTCCCAATGACGAGGTGATGAACCAAA	1059
QY	340	snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG	360
DB	1060	ATTTGGCTATTATGACGCTATGCTTGGAGAAGAACACACAGATCCATCGGCCCGCGTG	1119
QY	360	luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaT	380
DB	1120	AGAGTCCCAAGGAGTACCGACAGCGAGCCCTACTGGAAGAAAGAACTGCTTTTCTCGCTT	1179
QY	380	YrAspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIleP	400
DB	1180	ATGATGTTTTTGAATTCCTTTTGTGGATCCGATTCATGGACTCCAGAGAAGTGAATC	1239
QY	400	roLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG	420
DB	1240	CCAAAGAGATTGCAAGAGAAACAGAGTCAGAACCGGAAACAGCCGTACGCCATCTCCCAGG	1299
QY	420	luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL	440
DB	1300	AGATTGGGAACCTTATGAAGGAATCGAGACCTTGTGGAAGAGAAGACCAAGAGTCA	1359
QY	440	euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuT	460
DB	1360	TGGATGTGAGCAGACTGACCGGGAAGTGGCCCTGCTGTATGAAGGCACTAGTCTCA	1419
QY	460	hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer	479
DB	1420	CCATGAATCCAACTCTCTGAATGGTTC-CCAGCGGGTGGTGTGAGCGCGGTAACTCTCT	1478
QY	480	AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly	499
DB	1479	GACCAACGAGTGTGAGGAGCTGACAGAGCTGACCAATGTGGCAGCAACCTCAGGAGATGC	1538
QY	500	TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe	519
DB	1539	TACCGGGGTGAGACCTCCCCACATCTCCCAATGAAAGTTCTATGGTGTCACTGTCTTC	1598
QY	520	LysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyr	539
DB	1599	AAAGCCCTCAAGCTGGGGCAAGAGGCAAGTCTCTCTGACAGAGTGCCCACTGTACTAC	1658
QY	540	AsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeu	559
DB	1659	AACGTGACGGAAGAGTGGCGGCGCATCATGAGTCCCTACTTCCGCTGGATACGCCCTC	1718
QY	560	TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArg	579
DB	1719	TACTTTTCTCTACTCTCATCTGCTGTCGCGCACTGCCATCGAAGAGGTCCAGGACAGAGG	1778
QY	580	LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal	599
DB	1779	AAGGATGATGATGATCATCCAGTCCAGTCCAGCAACTGCAATCTCTGAATGCCGAGACCTCG	1838
QY	600	CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn	619
DB	1839	TGTGTCAAAGAGCCCGACGCTACACCTTCCGCGACTACAGCGCCATCTTTTACCTTAAAT	1898
QY	620	GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThAla	639
DB	1899	GGGGACTTCGATGGCGGAACTTTTATTCTACTGAACTGATGTCGCAAGACCGTGACGGA	1958
QY	640	GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis	659





Db 1931 AGTCCAGCAGAGAGAGGATGATGATCATCCAGTCCACGCTGGACAACTGCATCTCGA 1990  
QY 594 snAlaGluThrLeuValCysValLysGluProAlaTyrThrPheArgAspTyrSerA 614  
Db 1991 ATGCCGAGACCCCTCGTGTGTGTCAAAGAGAGCCGCCAGCCCTACACCTTCGCGACTACAGCG 2050  
QY 614 laileLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspA 634  
Db 2051 CCATCTCTTACCTAATGGGACCTTCGATGGCGGAACTTTTATTCTAGTAATGGATG 2110  
QY 634 laLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPheSerSerg 654  
Db 2111 CCAAGACCGTGCAGCAGAGGTGCAGCCTCAGTGTGGAAGAGCGGTGGATTCTCTTCAG 2170  
QY 654 lyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaL 674  
Db 2171 GCACCTGAAACCCCATGGAGTGAAGGTGTACCCAGGGGCGAGCGTGTGCCATCGCCC 2230  
QY 674 euTrpPheThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspAspLeuV 694  
Db 2231 TGTGTTTCACTGACCTCGACACAGCAGCGGAGCGGAGCGGTGCAGGCAGATGACCTGG 2290  
QY 694 alLysMetLeuPheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlag 714  
Db 2291 TGAAGATCTCTTACGCCCGCAGAAGAGATGGACCTCTCCAGAGCAGCGCCCTGGATGCC 2350  
QY 714 lngGlnGlyProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProlLysA 734  
Db 2351 AGCAGGGCCCCCGAACCTGCACAAGAGTCTCTCTCAGGCGATGCAATCGAAGCCCAAGG 2410  
QY 734 spGluLeu 736  
Db 2411 ATGAGCTA 2418

## RESULT 6

US-09-728-952-51  
; Sequence 51, Application US/09728952  
; Patent No. US2002011302A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Zhou, Ping  
; APPLICANT: Goodrich, Ryle  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Yamazaki, Vicki  
; APPLICANT: Ujwal, Manusha L.  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. US2002011302A1el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 799  
; CURRENT APPLICATION NUMBER: US/09/728,952  
; CURRENT FILING DATE: 2000-11-30  
; NUMBER OF SEQ ID NOS: 101  
; SOFTWARE: pt\_FL\_genes Version 2.0  
; SEQ ID NO 51  
; LENGTH: 2753  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(2418)  
US-09-728-952-51

## Alignment Scores:

Pred. No.:	0	Length:	2753
Score:	380.00	Matches:	580
Percent Similarity:	99.32%	Conservative:	0
Best Local Similarity:	99.32%	Mismatches:	2
Query Match:	51.63%	Indels:	4
DB:	9	Gaps:	0

US-10-045-815-4 (1-736) x US-09-728-952-51 (1-2753)

QY 155 LysIleAsnLysLeuGluLysAlaValAlaAlaHisThrPhePheValGlyAsnPro 174  
Db 673 AGATCAACAAGTTGGAGAAAGCTGTGTCTGCAGCACACACCTTCTTCGTGGCAATCCT 732  
QY 175 GluHisMetGluMetGlnGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGlu 194  
Db 733 GAGCACAATGGAATGAGCAGAACCTTAGACTATTACCAAAACCATCTCTGGAGTGAAGGAG 792  
QY 195 AlaAspPheLysAspLeuGluThrGlnProHisMetGlnGlnPheArgLeuGlyValArg 214  
Db 793 GCCGACTTCAAGGATCTTGAGACTCAACCCCATATGCAAGATTTTCGACTGGAGTGGGA 852  
QY 215 LeuTyrSerGluGluGlnProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGlu 234  
Db 853 CTCTACTCTAGAGGAAACAGCCACAGGAAGCTGTGCCCCACCTAGAGCGCGCGCTGCAAGAA 912  
QY 235 TyrPheValAlaTyrGluGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGly 254  
Db 913 TACTTTGTGCCCTATGAGAGGTGCCGTGCCCTCTCGAAGAGGCCCTATGACTACATGCG 972  
QY 255 TyrAsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGln 274  
Db 973 TACAACCTACTTGTAGTACAAACGCTGACCTTCCAGGCCATCACAGATCATTATACATCCAG 1032  
QY 275 ValLeuAsnCysLysGlnAsnCysValThrGluLeuAlaSerHisProSerArgGluLys 294  
Db 1033 GTCCCTCAACTGTAAAGCAAGACTGTGTCCAGGAGCTGTCTCCCAAGTTCGAGAGAG 1092  
QY 295 PropheGluAspPheLeuProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIle 314  
Db 1093 CCCTTTGAAGACTTCTCCCATCGCATTAATATATCTGCAGTTTGCCTACTATAACATT 1152  
QY 315 GlyAsnTyrThrGlnAlaGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAs 334  
Db 1153 GGGAAATTATACACAGGCTGT-TGAATGTGCCAAGACCTATCTCTCTTCTCCCAATGA 1211  
QY 334 pGluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrAr 354  
Db 1212 CGAGGTGATGAACCAAAATTTGGCTTATATGCAGCTATGCTTGGAGNAGAACACCCAG 1271  
QY 354 gSerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysG 374  
Db 1272 ATCCATCGGCCCCCGTGAGAGTGGCCAGAGTAGCCACAGCGAAGCTACTTGGAAAAAGA 1331  
QY 374 uLeuLeuPhePheAlaTyrAspValPheGlyIleProPheValAspProAspSerTrpTh 394  
Db 1332 ACTGCTTTTCTTCGCTTATGATGTTTTTGGAAATTTCCCTTTGTGGATCCGATTCATGGAC 1391  
QY 394 rProGluValIleProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAl 414  
Db 1392 TCCAGAGAAGTGAATTCCTCAAGAGATTGCAAGAGAAACAGAAAGTCAGAACGGGAAACAGC 1451  
QY 414 aValArgIleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThrLeuValGluG 434  
Db 1452 COTACGCATCTCCAGGAGATTGGAACTTATGAAGAAATCGAGACCTTTGTGGAGAA 1511  
QY 434 uLysThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTy 454  
Db 1512 GAAGACCAAGGAGTCACTGGATGTGAGCAGACTGACCCGGGAAGGTGGCCCCCTGCTGTA 1571  
QY 454 rGluGlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValM 474  
Db 1572 TGAAGGCATCAGTCTCACCATGAATCAAACTCCTCAATGGTTTC-CCAGCGGTGTGTA 1630  
QY 474 etAspGlyValIleSerAspHisGluCysGlnGlnLeuGlnArgLeuThrAsnValAlaA 494  
Db 1631 TGGACGGCGTAAATCTCTGACCACCGAGTGTCAAGAGCTGCAGAGACTGACCAATGTGCGAG 1690  
QY 494 laThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheT 514  
Db 1691 CAACCTCAGGAGATGGCTACCGGGTTCAGACCTCCCAACATCTCCCAATGAAAGTTCT 1750





QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
Db 952 CCATCCCATTAATTAATTAATTCGAGTTGGCTACTATACATTGGGAATTAACAAAGCT 1011  
QY 321 GlyGluCysAlaIysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
Db 1012 GGTGAATGTGCCAAGACCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1071  
QY 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360  
Db 1072 TTGGCCTATTATGCAGCTATGTTGAGAAAGAACACACAGATCCATCGGCCCGCTGAG 1131

## RESULT 8

US-10-257-174-1  
; Sequence 1, Application US/10257174  
; Publication No. US20040034194A1  
; GENERAL INFORMATION:  
; APPLICANT: Agarwal, Pankaj  
; APPLICANT: Murdock, Paul R.  
; APPLICANT: Rizvi, Safia K.  
; APPLICANT: Smith, Randall F.  
; APPLICANT: Xiang, Zhaoqing  
; TITLE OF INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GP50022  
; CURRENT APPLICATION NUMBER: US/10/257,174  
; CURRENT FILING DATE: 2002-10-10  
; PRIOR APPLICATION NUMBER: PCT/US01/11797  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/196,603  
; PRIOR FILING DATE: 2000-04-13  
; PRIOR APPLICATION NUMBER: 60/199,417  
; PRIOR FILING DATE: 2000-04-24  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-257-174-1

## Alignment Scores:

Pred. No.: 5,156-274 Length: 2127  
Score: 275.00 Matches: 375  
Percent Similarity: 99.47% Conservative: 0  
Best Local Similarity: 99.47% Mismatches: 1  
Query Match: 37.36% Indels: 2  
DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x US-10-257-174-1 (1-2127)

QY 361 SerAlaIysGluTyrArgGlnArgSerLeuLeuGlyGluLeuLeuPhePheAlaTyr 380  
Db 997 AGTGCCAAAGAGTACCCGACGGAAGCTACTGGAAAAAGAACTGCTTTTCTTCGCTTAT 1056  
QY 381 AspValPheGlyIleProPheValAspProAspSerTyrProGluGluValIlePro 400  
Db 1057 GATGTTTTTGAATTCCTTTGTGATCCGATTCAGGATTCAGGATTCAGGATTCAGGATTC 1116  
QY 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
Db 1117 AAGAGATTGCAAGAGAACAGAACTCAGAACCGGAAACAGCCGATCAGTCTCCCAAGGAG 1176  
QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440  
Db 1177 ATTGGGAACCTTATGAAGGAATCGACCTTGTGGAAGAGAACCAAGGATTCAGTCT 1236  
QY 441 AspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
Db 1237 GATGTGACGAGACTGACCCGGAAGGTGGCCCTGCTGTATGAAGGCATCAGTCTCACC 1296  
QY 461 MetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSerAs 480  
Db 1297 ATGAACCTCCAACTCTCTGAATGGTTC-CCAGCGGGTGGTATGACGGCGGTATCTCTGA 1355

QY 480 pHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
Db 1356 CCACGAGTGTGAGAGCTGCAGAGACTGCACCAATGTGGCAGCAACCTCAGGAGATGGCTA 1415  
QY 500 rArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheIy 520  
Db 1416 CCGGGGTGACAGCTCCCACTACTCCCAATGAAAAGTTCTATGGGTGTCACGTCTTCAA 1475  
QY 520 sAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAs 540  
Db 1476 AGCCCTCAAGCTGGGCAAGAGCAAGTTCTCTTCAGAGGTGCCCACTGCTACTACAA 1535  
QY 540 nValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
Db 1536 CGTGACGGAGAGGTGCGCGCATCATGGAGTCTTCTCCGCTGGATACGCCCTCTTA 1595  
QY 560 rPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgly 580  
Db 1596 CTTTTCTACTCTCACTCTGTTGCGGCACCTGCCATCGAAGAGGTCCAGGACAGAGAA 1655  
QY 580 sAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCy 600  
Db 1656 GGATGATAGTCATCCAGTCCAGTGGACAACTGCATCTGAAATGCGGAGACCTCGTGTG 1715  
QY 600 sValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGl 620  
Db 1716 TGTCAAGAGAGCCCGAGCCCTACACCTTCGCGACCTACAGCGCCATCTTTACTTAATGG 1775  
QY 620 yAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaIysThrValThrAlaGl 640  
Db 1776 GGACTTCGATGCGGAAACTTTTATTTTCACTGAATGGAATGCAAGACCGTGCAGGAGA 1835  
QY 640 uValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGl 660  
Db 1836 GGTGACGCTCAGTGTGGAAGAGCCGTGGATCTCTTCAGGACATGAAACCCACATGG 1895  
QY 660 yValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAspPr 680  
Db 1896 AGTGAAGGCTGTCAACAGGGGCGACGCTGTGCCATCGCCCTGTGTTCACTCCCTGACCC 1955  
QY 680 oArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPr 700  
Db 1956 TCGACACAGCAGCGGGACAGGTCAGGACAGATCACTGTGTGAAGATGCTCTTTCAGGCC 2015  
QY 700 oGluGluMetAspLeuSerGlnGluProLeuAspAlaGlnGlnGlyProProGluPr 720  
Db 2016 AGAAGAGATGACCTCTCCAGGACGACCCCTGATGCCAGGAGGTGCCCGCCCAACC 2075  
QY 720 oAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2076 TGCACAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2124

## RESULT 9

US-10-302-172-227  
; Sequence 227, Application US/10302172  
; Publication No. US20040053250A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Drmanac, Radjoje T.  
; TITLE OF INVENTION: No. US20040053250A1el Arginine-rich Protein-like Nucleic Acids and  
; FILE REFERENCE: 803 1NCP  
; CURRENT APPLICATION NUMBER: US/10/302,172  
; CURRENT FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 10/225,251  
; PRIOR FILING DATE: 2002-08-20  
; PRIOR APPLICATION NUMBER: PCT US02/05095  
; PRIOR FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: US 09/799,451  
; PRIOR FILING DATE: 2001-03-05  
; NUMBER OF SEQ ID NOS: 950

```

; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 227
; LENGTH: 2152
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (32)..(2149)
US-10-302-172-227

Alignment Scores:
Pred. No.: 6,14e-242 Length: 2152
Score: 244.00 Matches: 358
Percent Similarity: 98.90% Conservative: 0
Best Local Similarity: 98.90% Mismatches: 2
Indels: 4
Query Match: 33.15%
DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x US-10-302-172-227 (1-2152)

QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValValAlaAlaSer 20
Db 32 ATGGCGGTAGCGGCTTGAAGCTGCTGACACACATGCTGGCTGCGTGGCGCTGGCTCC 91
QY 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla 40
Db 92 CAAGCCAGGTGCGAGTCCGAGGCGAGGATGGGCGATGGTACGCTGATCTGCTTCGCC 151
QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60
Db 152 GAGGGGACCGCAGCTACGGCGCGGGGACTGGCGCGGGTGGTCTCTGAGCATGGAAACGG 211
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
Db 212 GCGTGGCGCTCCCGGGGAGCCCTCCGGCCCTTCGCTGGCTGGCGCCACCGAGTGGCC 271
QY 81 AlaAspPheProTyrGluLeuAspProAspTyrSerProSerProAlaGlnAlaSerGly 100
Db 272 GCCGACTTCGCTGGGAGTGGACCCCGACTGGTCCCGCCAGCGCGGCGCCAGCGCTCGGC 331
QY 101 AlaGly-AlaLeuArgAspLeuSerPheGlyGlyLeuLeuArgAlaAlaCysLe 120
Db 332 GCCGC-CCGCCCTGGCGACCTGAGCTTCTTCGGGGGCTTCTGGCTGCGCTGGCTGCC 390
QY 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140
Db 391 CGCGCGCTGCTCGGCGCGCGCGCGCCACTCGCTCAGCGAGAGATGGAGTGGAGTT 450
QY 140 eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysLeuAsnLysLeuG 160
Db 451 CCGCAAGCGGAGCGCCCTACAACTACCTGAGCTGCGCTTCTCAAGATCAACAAGTTGGA 510
QY 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyValAsnProGluHisMetGluMetG 180
Db 511 GAAAGCTGTTGCTGAGCAGCACACACTTCTTCTGGGCAATCTGAGCACATGGAAATGCA 570
QY 180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200
Db 571 GCAGAACCTAGACTATTACCAAAACCATCTCTGGAGTGAAGGAGCGCGACTTCAAGATCT 630
QY 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGlu 220
Db 631 TGAGACTCAACCCCATATTCGAAGAAATTTGAGATGGAGTGGACTCTACTCAGAGGAACA 690
QY 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrG 240
Db 691 GCCACAGAGAGTGTGGCCACCTAGAGCGCGCTGCAAGAATACTTTTGGGCTATGA 750
QY 240 uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTy 260
Db 751 GGAGTGGCGTGGCTCTGGCAGGGGCCCTATGACTACGATGGTCAACTACCTTGGATGA 810
QY 260 rAsnAlaAspLeuPheGlnAlaAlaThrAspHisTyrIleGlnValLeuAsnCysLysG 280
```

```

Db 811 CAACGCTGACCTCTTCCAGGCCATCACAGATCAATTACATCCAGTCTCAACTGTAAGCA 870
QY 280 rAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300
Db 871 GAACCTGTGTACGGAGCTTGTCTCCACCCCAAGTCGAGAGAAGCCCTTTGAAGACTTCTCT 930
QY 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320
Db 931 CCCATCGCATTTATATTATCTGCGAGTTGCGCTACTATAACATTGGGAATTATACACAGC 990
QY 320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340
Db 991 TGT-TGAATGTGCCAAGACCTATCTTCTCTTCTCCCAATGACAGGTGATGACACCAA 1049
QY 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArg 360
Db 1050 ATTTGGCTTATTATGCGATGCTGTTGGAGAAGAACACACAGATCCATCGGCCCTCGTG 1109
QY 360 lu 360
Db 1110 AG 1111

RESULT 10
US-10-302-172-226
; Sequence 226, Application US/10302172
; Publication No. US20040053250A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20040053250A1el Arginine-rich Protein-like Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 803 1CNCF
; CURRENT APPLICATION NUMBER: US/10/302,172
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/225,251
; PRIOR FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: PCT US02/05095
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 09/799,451
; PRIOR FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 950
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 226
; LENGTH: 1437
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (32)..(1255)
US-10-302-172-226

Alignment Scores:
Pred. No.: 3,13e-216 Length: 1437
Score: 219.00 Matches: 406
Percent Similarity: 99.02% Conservative: 0
Best Local Similarity: 99.02% Mismatches: 2
Query Match: 29.76%
DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x US-10-302-172-226 (1-1437)

QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValValAlaAlaSer 20
Db 32 ATGGCGGTAGCGGCTTGAAGCTGCTGACACACATGCTGGCTGCGTGGCGCTGGCTCC 91
QY 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla 40
Db 92 CAAGCCAGGTGCGAGTCCGAGGCGAGGATGGGCGATGGTACGCTGATCTGCTTCGCC 151
QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60
```

Db 152 GAGGGACCGCAGCCTACGCGCGCGGAGCTGGCCCGGGTGGTCTCTGAGCATGGAACGG 211  
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
Db 212 GCGCTCGGCTCGGCGCAGCCTTCGCGCCCTTCGCTGCGTGCAGCAGCAGGTGCTCC 271  
QY 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
Db 272 GCGCAGCTCCGCTGGAGCTGACCCGAGCTGGTCCCGCAGCCGCGCCAGGCTCGGCG 331  
QY 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyLeuLeuArgAlaAlaCysLe 120  
Db 332 GCGCG-CGCGCTCGGCGCAGCTGAGCTTCCTCGGGGGCTTCGCTGCGCTGCGCTGCT 390  
QY 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140  
Db 391 GCGCGCTGCTCGGCGCGCGCGCCGCTCGCTCAGCGAGAGATGGAGCTGGAGTT 450  
QY 140 eArgLysArgSerProTrpAsnTyrluGlnValAlaTyrluPheLysIleAsnLysLeuGl 160  
Db 451 CCGCAAGCGGAGCCCTACACTACCTGAGCTGCGCTACTTCAAGATCAACAAGTTGGA 510  
QY 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGl 180  
Db 511 GAAAGCTGTTGCTGAGCAGCACACCTTCTTCTGGGCAATCTGAGCACATGGAAATGCA 570  
QY 180 nGlnAsnLeuAspTyrluGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
Db 571 GCAGAACCTAGACTATTACCAAAACCATGCTGGAGTGAAGGAGCGCGACTTCAAGGATCT 630  
QY 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrluGluGl 220  
Db 631 TGAGACTCAACCCCATATGCAAGAAATTCGACTGGAGTGGAGTCTACTCAGAGGAACA 690  
QY 220 nProGlnAlaValProHisLeuGluAlaLeuGlnGluTyrluPheValAlaTyrluGl 240  
Db 691 GCCACAGGAAGCTGTGCCCCACCTAGAGCGCGCTGCAAGAAATCTTGTGGCTATGA 750  
QY 240 uGluCysArgAlaLeuGlyGluGlyProTyrluAspTyrluAspGlyTyrluGluTyrlu 260  
Db 751 GGAGTGGCTGCTGCGAAGGCGCCATGACTAGCTAGCTGCTGCTGCTGCTGCTGCTGCT 810  
QY 260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrluGlnValLeuAsnCysLysGl 280  
Db 811 CAACGCTGCTCTCCAGGCGCATCAGATCATCATCAGGCTCTCACTGTAAGCA 870  
QY 280 nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
Db 871 GAACGTGTCAAGGAGCTTGTCTCCCAAGTCGAGAGAAGCCCTTTGAAGACTTCCT 930  
QY 300 uProSerHisTyrluAsnTyrluGlnPheAlaTyrluAsnIleGlyAsnTyrluGlnAl 320  
Db 931 CCCATGCAATTAATTAATCTGCAAGTTTGCTACTATAACATTGGGAATTAACAGGC 990  
QY 320 aGly-GluCysAlaLysThrTyrluLeuPhePheProAsnAspGluValMetAsnGlnA 340  
Db 991 TGT-TGAATGTGCAAGACCTATCTTCTCTCTTCTCCCAATGACAGGTGATGAACAAA 1049  
QY 340 snLeuAlaTyrluAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG 360  
Db 1050 ATTGGGCTATTATGCAAGCTATGCTTGGAGAAGAACACACCATCATCGGCCCGCTG 1109  
QY 360 luSerAlaLysGluTyrluArgGlnArgSerLeuLeuGluLysGluLeuPhePheAlaAt 380  
Db 1110 AGAGTGCAGGAGTACCGACAGCGAGCTACTGGAAAAGAACTGCTTTCTTCGCTT 1169  
QY 380 yrAspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIleP 400  
Db 1170 ATGATGTTTGGAAATTCCTTTGTGATCCGATTCTGACTCCAGAGAAGTGAATTC 1229  
QY 400 roLysArgLeuGlnGluLysGlnLys 408  
Db 1230 CCAAGAGATTGCAAGAGAAACAGAG 1255

## RESULT 11

US-10-085-783A-6419  
; Sequence 6419, Application US/10085783A  
; Publication No. US20040037841A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: Liew, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2002  
; CURRENT APPLICATION NUMBER: US/10/085,783A  
; CURRENT FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/275,017  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 58994  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 6419  
; LENGTH: 337  
; TYPE: DNA  
; ORGANISM: Human  
US-10-085-783A-6419

Alignment Scores:  
Pred. No.: 4,61e-105 Length: 337  
Score: 111.00 Matches: 111  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 15.08% Indels: 0  
DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x US-10-085-783A-6419 (1-337)

QY 619 AsnGlyAspPheAspGlyGlyAsnPheTyrluLeuAspAlaLysThrValThr 638  
Db 3 AATGGGAGCTTCGATGGCGGAACTTTTATTCTCACTGAACCTGGATGCCAAGACCGTGACG 62  
QY 639 AlaGluValGlnProGlnCysGlyValArgAlaValAlaGlyPheSerGlyTyrluGluAsnPro 658  
Db 63 GCAGAGGTGCAGCCTCAGTGTGGAAGAGCCGTGGGATTTCTTTCAAGGACCTGAAACCCA 122  
QY 659 HisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeu 678  
Db 123 CATGAGTGAAGCTGTCAACAGGGGGCAGCGCTGTGCCATCGCCCTGTGGTTCAACCTG 182  
QY 679 AspProArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMetLeuPhe 698  
Db 183 GACCTTCGACACAGCAGCGGAGCAGGAGTGGAGGATGACCTGGTGAAGATGCTCTTC 242  
QY 699 SerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProPro 718  
Db 243 AGCCCAAGAGATGAGCTCTCCAGGAGCAGCCCTGGATGCCAGAGGGCCCCCCCC 302  
QY 719 GluProAlaGlnGluSerLeuSerGlySerGlu 729  
Db 303 GAACCTGCACAGAGTCTCTCTCAGGCGAGTGAA 335

## RESULT 12

US-10-242-535A-6419  
; Sequence 6419, Application US/10242535A  
; Publication No. US20040013663A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: Liew, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2005  
; CURRENT APPLICATION NUMBER: US/10/242,535A  
; CURRENT FILING DATE: 2002-09-12  
; PRIOR APPLICATION NUMBER: US 10/085,783  
; PRIOR FILING DATE: 2002-02-28



```
DB: 14 Gaps: 0
US-10-045-815-4 (1-736) x US-10-045-815-5 (1-2416)
QY 612 TyrSerAlaIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGlu 631
Db 1851 TRACGGCCATCCCTTACTCAATGGCAGCTTCGATGGAGAACTTTTACTTCACAGAA 1910
QY 632 LeuAspAlaLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPhe 651
Db 1911 CTAGATGCCAAGACATGTGACGCGACAGGTGCAGCCCCAGTGTGGAAGGCTGTGGGATTC 1970
QY 652 SerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAla 671
Db 1971 TCTTCTGCACCTGAGAACCCACATGGAGTGAAGGCTGTCCACGAGGGGCGACGCTGGGCC 2030
QY 672 IleAlaLeuTppPheThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAsp 691
Db 2031 ATCGCCCTGTGTTTACGCTGGATCCTCGGCACAGTGNAGAGAGACAGGGTGCAGGCAGAT 2090
QY 692 AspLeuValLysMetLeuPheSerProGluGlu 702
Db 2091 GACCTGTGAAGATGCTGTTCCAGCCAGAGAG 2123

RESULT 15
US-09-918-995-29327
; Sequence 29327, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29327
; LENGTH: 465
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(465)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-29327

Alignment Scores:
Pred. NO.: 4.16e-80 Length: 465
Score: 87.00 Matches: 87
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 11.82% Indels: 0
DB: 10 Gaps: 0

US-10-045-815-4 (1-736) x US-09-918-995-29327 (1-465)
QY 650 GlyPheSerSerGlyThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArg 669
Db 113 GGATTCCTTCAGCAGCTGAAACCCACATGGAGTGAAGGCTGTCCACGAGGGGGCAGCGC 172
QY 670 CysAlaIleAlaLeuThrPheThrLeuAspProArgHisSerGluArgAspArgValGln 689
Db 173 TGTGCCATCGCCCTGTGGTTTCCCTCGACCCCTCGACACAGCGAGCGGACAGGGTGCAG 232
QY 690 AlaAspAspLeuValLysMetLeuPheSerProGluGluMetAspLeuSerGlnGluGln 709
Db 233 GCAGATACCTGGTGAAGATGCTCTTCAGCCCGAGAGAGATGGACCTCTCCAGGAGCAG 292
QY 710 ProLeuAspAlaGlnGlnGlyProProGluProAlaGlnGluSerLeuSerGlySerGlu 729
```

```
Db 293 CCCTCGATGCCCGCAGCGGGCCCCCGGAACTGTCACAGAGAGTCTCTCTCAGGCACTGAA 352
QY 730 SerLysProLysAspGluLeu 736
Db 353 TCGAAGCCCCAAGGATGAGCTA 373

Search completed: July 18, 2004, 23:47:46
Job time : 1127 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 10:00:30 ; Search time 6548 Seconds  
(without alignments)  
3356.531 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 3870  
Sequence: 1 MAVRALKLTLLAVVAARS.....PPPAQESLSGSESKPKDEL 736

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DRV=xlp  
-Q/cgn2\_1/USPTO.spool\_p/US10045815/runat\_14072004\_123008\_16825/app.query.fasta\_1.903  
-DB=EST -Qfmt=fastap -SUFFIX=std.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCAIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10045815 @CGN 1 1 6425 @runat\_14072004\_123008\_16825 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEOQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pin:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_pbg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3819	98.7	2583	11	BC004160	Homo sapi
2	3379	87.3	3186	11	AK030436	Mus muscu
3	3377.5	87.3	2587	11	AK010578	Mus muscu
4	2574	66.5	2211	29	AY411090	Homo sapi
5	2370	61.2	1836	29	AY411092	Mus muscu
6	1685	43.5	1746	29	AY411091	Pan trogl
7	1623	41.9	999	14	CA488534	AGENCOURT
8	1532	39.6	899	13	BU148566	AGENCOURT
9	1510.5	39.0	1101	13	EX364036	EX364036
10	1505	38.9	881	13	EU543499	AGENCOURT
11	1497	38.7	1108	12	EM473346	AGENCOURT
12	1473.5	38.1	1201	13	EX334235	EX334235
13	1459.5	37.7	1201	13	EX442345	EX442345
14	1455	37.6	926	9	AL521774	AL521774
15	1412	36.5	913	12	BG679334	BG679334
16	1392	36.0	905	13	BU147415	AGENCOURT
17	1387	35.8	848	13	BU177586	AGENCOURT
18	1360.5	35.2	902	13	EX392014	EX392014
19	1331	34.4	899	13	BU179488	AGENCOURT
20	1321.5	34.1	955	13	BQ678346	AGENCOURT
21	1319	34.1	870	13	BQ676738	AGENCOURT
22	1319	34.1	889	13	BQ678884	AGENCOURT
23	1317.5	34.0	867	14	CF265106	AGENCOURT
24	1298	33.5	939	13	EX353083	EX353083
25	1293	33.4	1201	13	EX445087	EX445087
26	1281	33.1	766	12	BI856252	BI856252
27	1276	33.0	929	13	BQ677232	AGENCOURT
28	1264.5	32.7	991	13	BX380701	BX380701
29	1253	32.4	884	13	BU190877	AGENCOURT
30	1253	32.4	915	14	CB196719	AGENCOURT
31	1246.5	32.2	929	13	BQ722593	AGENCOURT
32	1245.5	32.2	941	13	BUB38872	AGENCOURT
33	1244.5	32.2	859	13	BQ213919	AGENCOURT
34	1234	31.9	767	12	BM041968	BM041968
35	1224.5	31.6	1192	13	BX398629	BX398629
36	1210	31.3	959	12	BM802225	AGENCOURT
37	1179.5	30.5	1656	29	AY412908	Homo sapi
38	1173	30.3	758	13	BQ178768	AGENCOURT
39	1169	30.2	981	13	BUB39422	AGENCOURT
40	1166	30.1	911	13	BQ720742	AGENCOURT
41	1160.5	30.0	705	10	BE294207	BE294207
42	1158.5	29.9	885	13	BQ680853	AGENCOURT
43	1158	29.9	991	13	BQ917813	AGENCOURT
44	1150	29.7	858	14	CD556679	AGENCOURT
45	1148	29.7	655	14	CB552308	MMSP0001_

# ALIGNMENTS

RESULT 1  
BC004160  
LOCUS BC004160  
DEFINITION Homo sapiens leucine proline-enriched proteoglycan (leprecan) 1,  
mRNA (CDNA clone IMAGE:2824480), containing frame-shift errors.  
ACCESSION BC004160  
VERSION BC004160.2 GI:37588922  
KEYWORDS HTC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2583)





QY 381 AspValPheGlyIleProPheValAspProAspSerThrProGluGluValIlePro 400  
| | | | |  
Db 1171 GATGTTTTTGGAAATCCCTTTGTGGATCCGAGTTCATGGACTCCAGAGAGATGATTCC 1230  
| | | | |  
QY 401 LysArgLeuGlnGluGlySerGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
| | | | |  
Db 1231 AAGAGATTGCAGAGAAACAGAGAGTCAAGCGGAAACAGCCGTACGATCTCCAGGAG 1290  
| | | | |  
QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluThrLysGluSerLeu 440  
| | | | |  
Db 1291 ATTGGGAACCTTATGAGGAAATCGAGACCTTTGTGGAAGAGAACCAAGAGAGTCACTG 1350  
| | | | |  
QY 441 AspValSerArgIleThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460  
| | | | |  
Db 1351 GATGTGAGCAGACTGACCCGGGAAGGTGGCCCTGCTGTATGAAGCATCAGTCTCAC 1410  
| | | | |  
QY 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480  
| | | | |  
Db 1411 ATGAACCTCCAACTCTCTGATGTTCCAGCGGTGTGATGACGCGTAACTCTGAC 1470  
| | | | |  
QY 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGlyTyr 500  
| | | | |  
Db 1471 CACGAGTGTGAGAGTGTGACAGACTGACCAATGTGCAGCAACCTCAGGAGATGGCTAC 1530  
| | | | |  
QY 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520  
| | | | |  
Db 1531 CGGGTTCAGACCTCCCACTATCTCCCATGAAAGTTCTATGCTGCTACTGCTTCAA 1590  
| | | | |  
QY 521 AlaLeuLysLeuGlnGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540  
| | | | |  
Db 1591 GCCTCAAGCTGGGGCAAGGCAAGGCAAGTTCCTCTGCAGAGTCCCACTGTACTACTAAC 1650  
| | | | |  
QY 541 ValThrGluLysValArgAlaGlyMetGluSerTyrPheArgLeuAspThrProLeuTyr 560  
| | | | |  
Db 1651 GTGACGAGAGAGTGGCGGCATCATGGAGTCTCTTCGCGCTGATAGCCCTCTAC 1710  
| | | | |  
QY 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580  
| | | | |  
Db 1711 TTTTCTTACTCTCATCTGGTGTGCGCACTGCCATCGAAGGTCAGGCGAGAGAGAA 1770  
| | | | |  
QY 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600  
| | | | |  
Db 1771 GATGATAGTATCATCGATCCAGTCCAGTGGCAACTGATCTGTAATCCGAGACCTCTGTGT 1830  
| | | | |  
QY 601 ValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620  
| | | | |  
Db 1831 GTCAAGAGCCCCAGGCTTACACTTCGCGACTACAGCCCATCTTTACTTAATGG 1890  
| | | | |  
QY 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640  
| | | | |  
Db 1891 GACTTCGATGGCGAAACTTTATTTTCACTGAATGGATGCCAAGACCCGTGACGGCAGAG 1950  
| | | | |  
QY 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660  
| | | | |  
Db 1951 GTGAGGCTCAGTGTGGAAGAGCGGTGGATTTCTTCAGGCACTGAAACCCACATGGA 2010  
| | | | |  
QY 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680  
| | | | |  
Db 2011 GTGAAGGCTGTCAACAGGGGGCAGCGCTGTGCCATCGCCCTGTGGTTCACCTGGACCT 2070  
| | | | |  
QY 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700  
| | | | |  
Db 2071 CGACACAGCGAGGGGACAGGGGTGAGGAGATGACCTGTGTGAAGATGCTCTTCAGCCCA 2130  
| | | | |  
QY 701 GluGluMetAspLeuSerGlnGluProLeuAspAlaGlnGlnGlyProProGluPro 720  
| | | | |  
Db 2131 GAAGAGATGACCTCTCCAGGAGCAGCCCTGGATGCCAGGCGGCCCCCGAACCT 2190  
| | | | |  
QY 721 AlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
| | | | |  
Db 2191 GCACAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2238  
| | | | |

RESULT 2  
AK030436  
LOCUS  
DEFINITION  
MUS MUSCULUS ADULT MALE PITUITARY cDNA, RIKEN FULL-LENGTH  
ENRICHED LIBRARY, CLONE:5330411L17 PRODUCT:LEPRECAN 1, FULL INSERT  
SEQUENCE.  
ACCESSION  
AK030436  
VERSION  
AK030436.1 GI:26326436  
KEYWORDS  
HTC; CAP TRAPPER.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
REFERENCE  
1 Carninci, P. and Hayashizaki, Y.  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
99279253  
REFERENCE  
2 Carninci, P. and Hayashizaki, Y.  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
20493374  
REFERENCE  
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
4 The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
20530913  
REFERENCE  
5 The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
6 Nature 409, 685-690 (2001)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
7 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
8 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
9 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
10 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
11 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
12 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
13 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
14 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
15 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
16 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
17 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
18 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
19 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
20 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
21 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
22 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
23 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
24 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
25 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
26 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
27 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
28 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
29 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
30 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
31 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
32 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
33 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
34 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
35 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
36 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
37 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
38 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
39 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
40 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
41 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
42 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
43 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
44 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
45 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
46 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
47 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
48 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
49 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
50 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
51 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
52 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
53 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
54 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
55 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
56 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
57 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
58 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
59 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
60 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
61 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
62 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
63 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
64 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
65 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
66 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
67 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
68 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
69 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
70 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
71 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
72 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
73 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
74 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
75 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
76 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
77 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
78 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
79 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
80 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
81 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
82 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
83 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
84 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
85 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
86 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
87 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
88 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
89 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
90 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
91 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
92 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
93 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
94 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
95 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
96 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
97 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
98 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
99 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
100 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
101 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
102 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
103 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
104 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
105 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
106 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
107 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
108 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
109 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
110 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
111 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
112 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
113 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
114 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
115 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
116 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
117 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
118 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
119 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
120 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
121 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
122 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
123 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
124 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
125 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
126 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
127 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
128 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
129 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
130 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
131 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
132 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
133 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
134 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
135 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
136 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
137 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
138 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
139 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
140 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
141 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
142 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
143 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
144 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
145 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
146 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
147 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
148 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
149 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
150 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
151 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
152 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
153 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
154 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
155 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
156 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
157 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
158 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
159 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
160 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
161 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
162 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
163 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
164 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
165 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
166 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
167 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
168 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
169 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
170 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
171 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
172 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
173 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
174 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
175 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
176 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
177 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
178 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
179 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
180 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
181 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
182 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
183 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
184 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
185 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
186 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861  
REFERENCE  
187 Nature 420, 563-573 (2002)  
TITLE  
JOURNAL  
MEDLINE  
PUBMED

Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN,  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: <http://genome.gsc.riken.go.jp/>  
URL: <http://fantom.gsc.riken.go.jp/>.

## FEATURES

## source

1..3186  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM.DB:533041117"  
/db\_xref="MGI:2333259"  
/db\_xref="taxon:10090"  
/clone="533041117"  
/sex="male"  
/tissue\_type="pituitary gland"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
44..2260  
/note="unnamed protein product; leprecan 1  
(MGI:MGI:1888921)  
putative"

## CDS

/codon\_start=1  
/protein\_id="BAC26962.1"  
/db\_xref="GI:26326437"  
/translation="MAVSERRLLAAMLAVALAALAAVAABEPGWDAAPDLIYABGT  
AAYSGDPGVNLNMRALRALRLRCCTRCATLPMAPDLIDLPDPSLODP  
GAALHLDRPFGAVLRPAALCLRLCPPSAHLSLELDLEFNKESYNYLQVAYFKIN  
KLEKAAAHITFVGNPERHMQNLDYITMSGVKREDFDLKAPMHFERLGVRL  
YSEKPEAVPHLEALAQEYVADBECLACGPDYDGYNDYLSADLUFQAITDHYV  
QVLNCKQNTLASHSPREKPEDELPSPHYNLYQFAYNIGNYQAIKCAKTYLLFF  
PNDVWHQNLAYITAMLEGEASISIPRENAEYRRRLLEKELFFAYDIFGIPFD  
PDSWTPEVLPKLOEKORSETAVASQIEIGLMKEIETLVEKTKESLDVSLTRPE  
GGPLLRYGILSTWNSKVLNGSORVMDGVISDDCELORLTNAATSGDGYRGOTSP  
HTNKEKYGVTVLKLALGQGVKVPLOQARMYNTKVKRVMEVFLDITPLYSYS  
HLVCRTALESQARKDSRPVHVDNCLINAEALMCIKEPPATPRDYSAIILYNGDF  
ROSERDRAVDLIVKMLFSPSEVDLPQEQPLPDQGSPEPGEESLDRSLHKLDEL"  
3172..3177  
/note="putative"  
3186  
/note="putative"

## polyA\_signal

## polyA\_site

## ORIGIN

## Alignment Scores:

Pred. No.: 0 Length: 3186  
Score: 3379.00 Matches: 644  
Percent Similarity: 91.75% Conservative: 34  
Best Local Similarity: 87.14% Mismatches: 57  
Query Match: 87.31% Indels: 4  
DB: 11 Gaps: 3

US-10-045-815-4 (1-736) x AK030436 (1-3186)

Qy 1 MetAlaValAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaAlaSer 20  
Db 44 ATGGCGGTGAGCGAAGAGGCTCTCGGAGCATGCTAGCGTGGCGCGCGCGCGCC 103  
Qy 21 ---GlnAlaGluValGluSerGluAlaGlyTTPGlyMetValThrProAspLeuLeuPhe 39  
Db 104 TTACGGGTTCGGCGGAGTCTGACCGGGATGGAGCTGGCAGCCCTGACCTGCTTTAC 163  
Qy 40 AlaGluGlyThrAlaAlaValAlaArgGlyAspTrpProGlyValValLeuSerMetGlu 59  
Db 164 GCAGGGGACCGCGGCTACTCGCGGGAGACTGGCGCGGGGTGGTCTCGAATGGAG 223  
Qy 60 ArgAlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCys 79  
Db 224 CGGCGCTCGGCTCGCGGGCGGCGCTCGTGCCTCGCGCTCGCGCACACACCTGT 283  
Qy 80 AlaAlaAspPheProTrpGluLeuAspProAspTrpSerProSerPro-----AlaGln 97

Db 284 GCCACCGAACTGCGGTGGGACCGGACCTGGATCTCGGTCCGGACCCCGAGCTGAGCCAG 343  
Qy 98 AlaSerGlyAlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAla 117  
Db 344 GACCCGGCGCGCGCGCGCTGACACACCTGCGCTTCTTCGAGCGCTGCTCGCGCGTGC 403  
Qy 118 AlaCysLeuArgGlyCysLeuGlyProAlaAlaHisSerLeuSerGluGluMetGlu 137  
Db 404 GCCTGCTTACCGCGCTGCTGCGCGCGCTCTGCGCACTTGTGAGTGAAGAACTGGAC 463  
Qy 138 LeuGluPheArgLysArgSerProTrpAsnTyrLeuGlnValAlaTyrPheLysIleAsn 157  
Db 464 CTGGAGTTCAACACAGCGGAGCCGCTACCACTACCTGACAGTGGCTTATTTCAAGATAAC 523  
Qy 158 LysLeuGluLysAlaValAlaAlaHisThrPhePheValGlyAsnProGluHisMet 177  
Db 524 AAGCTGGAGAAGCTGTGCTGCGGCACACACCTTCTTTTGGGCAATCTCTGACACATG 583  
Qy 178 GluMetGlnGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPhe 197  
Db 584 GAGATGCGCGAGAACCTCGACTATTACCAACCATGTCTGGGTGAAGAGGAGCAGACTTC 643  
Qy 198 LysAspLeuGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSer 217  
Db 644 AGGGATCTCGAGGCGCAAGCCCATATGATGAGTTTGGCTGGGGGTACGACTCTACTCA 703  
Qy 218 GluGluGlnProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheVal 237  
Db 704 GAGGAGAGCACACAGAGAGCTGTGCCACCTGGAGCGGACCTGCAAGAGTACTTTGTG 763  
Qy 238 AlaTyrGluGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyr 257  
Db 764 GCGGATGAGGAATGCGGTGCGCTCTGCGAAGGGCCCTATGACTACGACGGCTCAACATC 823  
Qy 258 LeuGluTyrAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsn 277  
Db 824 CTAGACTACGGCTGACCTCTTCAGGCCATCACAGTCAATTAGCTCCAGTCTCTCAAC 883  
Qy 278 CysLysGlnAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGlu 297  
Db 884 TGTAAGCAGAACTGTGTACGAGCTGCTCCACCCAGTAGGAGAAAGCCCTTTGAA 943  
Qy 298 AspPheLeuProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyr 317  
Db 944 GACTTCTCCCTTACACTATAATTACTACGTTTGCCTACTACACATTGGGAATAT 1003  
Qy 318 ThrGlnAlaGlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMet 337  
Db 1004 ACACAAGCTATTGAATGTGCCAAGACCTACCTCTCTCTTCTTCCCAATGATGAGGTGATG 1063  
Qy 338 AsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGly 357  
Db 1064 CACCAGAACTTGGCTTATTACAGCCCATGTTTGAGAAAGAGAGCCAGCTCCATCAGC 1123  
Qy 358 ProArgGluSerAlaLysGluTyrArgGlySerLeuLeuGluLysGluLeuLeuPhe 377  
Db 1124 CCGAGGAGATCGCGAGGATACCGAGCTCGAAGCTGTGGAGAAAGAACTGCTTTTC 1183  
Qy 378 PheAlaTyrAspValPheGlyIleProPheValAspProAspSerTrpThrProGluGlu 397  
Db 1184 TTCGTTATGACATTTTGGAAATTCCTTTGTGATCGGACTCATGAGTCCAGAGAG 1243  
Qy 398 ValIleProLysArgLeuGlnLysGlnLysSerGluArgGluThrAlaValArgIle 417  
Db 1244 GTGATTCACAGAGACTGCGAGAGAAACAGAGTCCGAAACCGGAAACCGCAGTCCGC--- 1300  
Qy 418 SerGlnGluIleGlyAsnLeuMetLysGluIleGluThrLeuValGluLysThrIys 437  
Db 1301 TCCAGGAGATGGCAACCTTATGAAGAGATCGAGACCTCTCGTGGAGGAGAACCAAG 1360  
Qy 438 GluSerLeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIle 457

Db	1361	GAGTCTCGATGTGAGCGGACTGACGGGAGAGTGGTCCCTCGTGTATGAAGGCATC	1420
Qy	458	SerLeuThrMetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyVal	477
Db	1421	AGTCTCACCATGAATTCCTCAAGTCTTGAATGGCTCCAGCGGGTGGTGATGGTGGTG	1480
Qy	478	IleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaIleSerGly	497
Db	1481	ATCTCTGATGATGAGTGGCAGAGCTGCAGAGACTGCACCAATGGCGGACCACTTCGGGA	1540
Qy	498	AspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThr	517
Db	1541	GATGGCTACCGAGTCCAGCTCCCCACACACCCCAATGAAGTTCTATGGTGTACT	1600
Qy	518	ValPheGlyAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeu	537
Db	1601	GTCTCAAGCTCTCAAGCTCGGAGAGGAAAGTTCCTCTGCAGTGGCCGCGATG	1660
Qy	538	TyrTyrAsnValThrGluLysValArgGileMetGluSerTyrPheArgLeuAspThr	557
Db	1661	TACTACAAGCTGACAGAGAGTGGGGCGTGCATGGAGTCTACTTCCGCTGGACAG	1720
Qy	558	ProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAla	577
Db	1721	CCCTCTATTCTCTATTCCCACTCGTGTGGCGCACTGCAATAGAAGAGTCCACGGCT	1780
Qy	578	GluAtrGlyAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThr	597
Db	1781	GAGAGAGGACGATGACCCCGTCCACGGTAACTGCATCTCTGTAATGCCGAAGCC	1840
Qy	598	LeuValCysValLysGluProProAlaTyrThrPheArgSpTyrSerAlaIleLeuTyr	617
Db	1841	CTCATGTGATCAAGGAGGCCCCAGCTACAGCTTCGCGGACTCAGCGCCATCTTTAC	1900
Qy	618	LeuAsnGlyAspPheAspGlyAsnPheTyrPheThrGluLeuAspAlaLysThrVal	637
Db	1901	CTCAATGGCGACTTCGATGGAGAACTTTTACTTTCACAGAATAGATGCCAAGCTGTG	1960
Qy	638	ThrAlaGluValGlnProGlnCysGlyArgAlaValAllyPheSerSerGlyThrGluAsn	657
Db	1961	ACGGCAGAGTGCAGCCCCAGTGTGGAGGGCTGTGGGATCTCTCTCGCATGAGAAC	2020
Qy	658	ProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThr	677
Db	2021	CCCATGTGGATGAAGCTGTCCACAGGGGAGCGCTCGCGCATCGCCCTGTGGTTCACG	2080
Qy	678	LeuAspProArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMetLeu	697
Db	2081	CTGGATCTCGCAAGTGTAGAGACACAGGGTGCAGCGCATGATGACCTGGTGAAGATCTG	2140
Qy	698	PheSerProGluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnIndIngPro	717
Db	2141	TTCAGCCCAAGAGGTGACCTCCCGAGGAACAGCCCTCGCTGCACAGCGGGCTCG	2200
Qy	718	ProGluProAlaGlnLysLeuSerGlySerGluSerLysProLysAspGluLeu	736
Db	2201	CCAGAGCTTCGAGAGAGTCTCTGTGACAGAGGGTCTCGTGCACAGGATGAGCTA	2257
RESULT	3		
LOCUS	AK010578	2587 bp mRNA linear HTC 20-SEP-2003	
DEFINITION		Mus musculus B5 cells cDNA, RIKEN full-length enriched library, clone:24i0024C15 product:16precin 1, full insert sequence.	
ACCESSION	AK010578		
VERSION	AK010578.1	GI:12846124	
KEYWORDS		HTC, CAP trapper	
SOURCE		Mus musculus (house mouse)	
ORGANISM		Mus musculus	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	
REFERENCE	1		
AUTHORS		Carninci, P. and Hayashizaki, Y.	
TITLE		High-efficiency full-length cDNA cloning	

```

/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:2410024C15"
/db_xref="MGI:1904948"
/db_xref="taxon:10090"
/clone="2410024C15"
/cell_type="ES cells"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
45..2273
/note="unnamed protein product; leprecan 1
(MGI|MGI:1888921)
putative"
/codon_start=1
/protein_id="BAB27041.1"
/db_xref="GI:12846125"
/translation="MAVSERRLLAAMLAVALAAALRAAESEFPGVDAVAPDLILYAEQT
AAVSRGDPGVNLMERALSRALRALRURCTRCAATELPWAPDLIDLPDFSLSDQP
GAALHDLRPFGLVLMERALSRALRALRURCTRCAATELPWAPDLIDLPDFSLSDQP
KLEKAAVAATFFVGNPEHMEMQNLDYOTMSGVKEADPRDLEAPMHEPRGLVRL
YSEKPOEAPHLEALAOEYEVADBECLCEGPDYDGVNLDVDSADLFOATIDHVV
QVLNCKONCVTELASHSRKPEDPELPSHYNLQFAYNIGNYTAISCAKTYLLFF
PNDVEMHONLAYTAMUGIEBEASISPRENAEYRRLRSLEKELLFFXADIFGIPVD
PDSMTPEVLPKRLQKQSERETAVRISQEIENLMBETLVEEKSLSDVSRUR
EGGPLLYGSLTWNLSVNGSORVMDGVISDECEQLORLTNAATSGDGYRGTS
PHTNEXFYGVTWLKALGOEGKVPLOSARMYNTVEKVRVMEYFRLDTPLYFSY
SHLCRTAIESQAEKDSHPVHVDNCLINAEALMCIKEPPAYTRDYSAILYNGD
PDGNYFTBEDATVTAEPQOCGRVGFSSGTENPHGVKAVTRGORCALALWFTLD
PRHSERDVAQDDLVKMLFSPBEVTSPPNSPCLTSRARQSLKSLQCTGRCRTRMSYD
NS"
polyA_signal 2565..2570
polyA_site 2587
/note="putative"
/origin="putative"
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 2587
Score: 3377.50 Matches: 646
Percent Similarity: 91.88% Conservative: 33
Best Local Similarity: 87.42% Mismatches: 57
Query Match: 87.27% Indels: 4
DB: 11 Gaps: 2
US-10-045-815-4 (1-736) x AK010578 (1-2587)
Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaLaser 20
Db 45 ATGCGGTGACGGAAGAGGCTGCTGGCAGCGATGCTAGCGGTGCGCGCGCGCGCGCC 104
Qy 21 ---GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPhe 39
Db 105 TTACGGGTGCGCGCGGAGTCTGACCGCGGATGGACGCGCCCTGACCTGCTTTAC 164
Qy 40 AlaGluGlyThrAlaAlaTyAlaArgGlyAspTrpProGlyValValLeuLeuSerMetGlu 59
Db 165 GCAGAGGGGACCGCGGCTACTCGCGCGGGAGCTGGCCCGGGGTGTCTTGACATGGAG 224
Qy 60 ArgAlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCys 79
Db 225 CGGCGGTGCGCTCGCGGGGCGGCTCGTGGCTGCGCTCGCGCTGCGGCTGCGGCTGTC 284
Qy 80 AlaAlaAspPheProTrpGluLeuAspProAspTrpSerProSerPro-----AlaGln 97
Db 285 GCCACCAACTGCGTGGGACCGGACCTGGATCTCGGTCCGACCCCGACCCCTGAGCCAG 344
Qy 98 AlaSerGlyAlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAla 117
Db 345 GACCCGGCGCGCGCGCCCTGCACGACCTGCGCTTCTTCGGAGCGGCTGCGCGCGTGC 404
Qy 118 AlaCysLeuArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluMetGlu 137
Db 405 GCCTGCTTACGCGCTGCTGCGGCGCGCCCTCTGCCCCACTTGTGAGTGAGAACTGGAC 464

```

```

Qy 138 LeuGluPheArgLysArgSerProTyRasnTyRasnLeuGlnValAlaTyRasnPheLysIleAsn 157
Db 465 CTGGAGTTCAACAGCGGAGCGGTACCACTACCTTGCAGGTGCGCTATTTCAAGATAAAC 524
Qy 158 LysLeuGluLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMet 177
Db 525 AAGCTGGAGAAAGCTGTGGCTGGGCAACACACCTTCCTTTGTGGCAATCTCTGACACATG 584
Qy 178 GluMetGlnGlnAsnLeuAspTyRasnTyRasnMetSerGlyValLysGluAlaAspPhe 197
Db 585 GAGATGCGGCAACCTCGACTATTACCAACCATGTCTGGGGTCAAGGAGGAGAGCTTC 644
Qy 198 LysAspLeuGluThrGlnProHisMetGlnGluPheArgLeuGluValArgLeuTySer 217
Db 645 AGGGATCTCGAGGCCAAGCCCATATGATGATGATTTGGCTGGGGGTACGACTCTATCTCA 704
Qy 218 GluGluGlnProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyPheVal 237
Db 705 GAGGAGAAAGCACAGAAAGCTGTGCCACCTGGAGGGGCGACTGCAAGAGTACTTTGTG 764
Qy 238 AlaTyRasnGluGluCysArgAlaLeuGluGlyProTyRasnTyRasnTyRasnTyR 257
Db 765 GCCGATGAGGAATGCGGTGCGCTCTGCGAAGGGCGCTTATGACTACGAGCGGTACAACTAC 824
Qy 258 LeuGluTyRasnAlaAspLeuPheGlnAlaIleThrAspHisTyRasnIleGlnValLeuAsn 277
Db 825 CTAGACTACAGCGCTGACCTCTTCCAGGCCATCACAGATCATTTAGCTCCAGTCTCTCAAC 884
Qy 278 CysLysGlnAsnCysValThrGluLeuAlaSerHisProSerArgGluLysPhePheGlu 297
Db 885 TGTAAGCAGAACTGTGTACGAGCTGGCTTCCACCAAGTAGGAGAAAGCCCTTTGAA 944
Qy 298 AspPheLeuProSerHisTyRasnTyRasnLeuGlnPheAlaTyRasnIleGlyAsnTyR 317
Db 945 GACTTCTCTCCCTCACACTATAATACCTACAGTTTGCTACTACAACTTGGGAACAT 1004
Qy 318 ThrGlnAlaGlyGluCysAlaLysThrTyRasnLeuPhePheProAsnAspGluValMet 337
Db 1005 ACACAAGCTATTGATGTGCCAAGACCTTACCTCTCTTCTTCCCAATGATGAGGTGATG 1064
Qy 338 AsnGlnAsnLeuAlaTyRasnAlaAlaMetLeuGluGluHisThrArgSerIleGly 357
Db 1065 CACCAAGATCTGGCTTATTACAGCCATGCTTGGAGAGAGAGGCCAGCTCCATCAGC 1124
Qy 358 ProArgGluSerAlaLysGluTyRasnArgGlnArgSerLeuLeuGluLysGluLeuPhe 377
Db 1125 CCGAGGAGAAATCGCAGGAATACCGAGCTCGAAGCCTGTGGAGAAAGAACTGCTTTTC 1184
Qy 378 PheAlaTyRasnValPheGlyIleProPheValAspProAspSerTrpThrProGluGlu 397
Db 1185 TTCGCTTATGACATTTTGGAAATTTCCCTTTGTGATCCGACTCCGAGCTCCAGAGAG 1244
Qy 398 ValIleProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIle 417
Db 1245 GTGATTTCCAAAGAGACTGCAGAGAAACAGAAAGTCGGAACCGCAACCGCAGTCCGCATC 1304
Qy 418 SerGlnGluIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLys 437
Db 1305 TCCCAAGGAGATGGCAACCTTATGAAGGAGATCAGACCCCTCGTGGAGGAGAAACCAAG 1364
Qy 438 GluSerLeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyRasnGluGly 457
Db 1365 GAGTCTTGGATGTGAGCCGACTGACCGGGAAGGTGGTCCCTGCTGATATAGGCGATC 1424
Qy 458 SerLeuThrMetAsnSerLysLeuLeuAsnGlyTyRasnArgValValMetAspGlyVal 477
Db 1425 AGTCTCACCAATGAATTCAAAAGTCTTGAATGGCTCCCGACGGGGTGGTGGATGGTGTG 1484
Qy 478 IleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGly 497
Db 1485 ATCTCTGATGATGAGTGCAGAGAGCTGCAGAGACTGACCAATGCGGAGCAACTTCGGA 1544
Qy 498 AspGlyTyRasnGlyGlnThrSerProHisThrProAsnGluLysPheTyRasnValThr 517

```

CDS



```
Qy 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260
Db 721 GAGTGGCGTGCCTCTCGAAGGGCCCTAGACTACGATGGCTACACTACCTTGAGTAC 780
Qy 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280
Db 781 AACCTGACCTCTCCAGGCCATCACAGATCATACATCCAGGTCTCACTGTAAGCAG 840
Qy 281 AsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300
Db 841 AACTGTGTTCACGAGCTGTCTCCACCAAGTCAGAGAAGCCCTTTGAAGACTTCCCTC 900
Qy 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320
Db 901 CCATCGCATATAATATCTGCAGTTTCCCTACTATTAACATTGGGAATATATACAGGCT 960
Qy 321 GlyGluCysAlaTyrThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340
Db 961 GTTGAATGTGCCAAGACTATCTTCTTCTCCCAATGACGAGGTGATGAACCAAAAT 1020
Qy 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgGlu 360
Db 1021 TTGGCTTATTATGAGCTATGCTTGGAGAAGAACACACAGATCCATCGGCCCCCGTGAG 1080
Qy 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaTyr 380
Db 1081 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1140
Qy 381 AspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIlePro 400
Db 1141 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1200
Qy 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420
Db 1201 AAGAGATTGCNAGAGAAACAGAGTCAAGACGGGAACACCGGTACGCATCTCCAGGAG 1260
Qy 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440
Db 1261 ATTGGGAACCTTATGAAGAAATCGACACCTTGTGGAGAGAAGACCAAGGAGTCACTG 1320
Qy 441 AspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuThr 460
Db 1321 GATGTGACGACACTGACCCGGGAAGTGGCCCTCTGCTGTATGAAGCATCAGTCTCACC 1380
Qy 461 MetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGlyValIleSerAsp 480
Db 1381 ATGAACCTCCAACTCCTGAATGGTTCCACGGGTGTGTGACGCGTAACTCTGTGAC 1440
Qy 481 HisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaIleThrSerGlyAspGlyTyr 500
Db 1441 CACGAGTGTGAGGAGTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGCTAC 1500
Qy 501 ArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPheLys 520
Db 1501 CGGGGTGAGACTCCCCACATCTCCCAATGAAGAAGTCTATGGTGTCACTGCTTCAAA 1560
Qy 521 AlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyrAsn 540
Db 1561 GCCCTCAAGCTGGGGCAAGAGGCAAAAGTTCCTCTGCAGAGTCCCACTGTACTACAAC 1620
Qy 541 ValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeuTyr 560
Db 1621 GTGACGAGAGAGTGGCGGCATCATAGACTCTACTTCCGCTGTGATACGCCCTCTTAC 1680
Qy 561 PheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArgLys 580
Db 1681 TTTTCTTACTCTCATCTGTGTGCGGCACCTGCCATCGCAAGAGGTCCAGGCGAGAGGAAG 1740
Qy 581 AspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuValCys 600
Db 1741 GATGATAGTATCATCGTCCACGAGGACACTGATCTCTGATCCGAGACCTTCGTGTGT 1800
Qy 601 ValLysGluProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsnGly 620
```

```
Db 1801 GTCAAGAGCCCCAGCCTACACCTTCGCGACTACAGCGCATCTTACTTAATGGG 1860
Qy 621 AspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAlaGlu 640
Db 1861 GACTTCGATGCGGAACCTTTTATTCTACTGAATGATGCAAGACCGTGACGNNNNN 1920
Qy 641 ValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHisGly 660
Db 1921 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1980
Qy 661 ValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAspPro 680
Db 1981 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2040
Qy 681 ArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSerPro 700
Db 2041 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2100
Qy 701 GluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProGluPro 720
Db 2101 GAAGAGATGGACCTCTCCAGAGCAGCCCTGATGCCAGCAGGGCCCCCGAACCT 2160
Qy 721 AlaGlnGluSerLeuSerGlySerGlySerLysProLysAspGluLeu 736
Db 2161 GCACAAGAGTCTCTCTCAGGCAGTGAATCGAAGCCCAAGGATGAGCTA 2208

RESULT 5
AY411092 1836 bp DNA linear GSS 16-DEC-2003
LOCUS Mus musculus LEPRE1 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY411092
VERSION AY411092.1 GI:39767060
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1836)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1836)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
Location/Qualifiers
source
1..1836
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>1836
/gene="LEPRE1"
/locus_tag="HCM4108"

gene

ORIGIN

Alignment Scores:
Pred. No.: 1,46e-241 Length: 1836
Score: 2370.00 Matches: 462
Percent Similarity: 82.10% Conservative: 15
Best Local Similarity: 79.52% Mismatches: 104
```

Query Match:	61.24%	Indels:	0
DB:	29	Gaps:	0
US-10-045-815-4 (1-736) x AY411092 (1-1836)			
Qy	156	IleAsnLysLeuGluLysAlaValAlaAlaAhiSThrPheValGlyAsnProGlu	175
Db	91	ATRAACAAGCTGGAGAAGCTGTGGCCGCGCACACACCTTCTTTGTGGCAATCTCTGAG	150
Qy	176	HisMetGluMetGlnGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAla	195
Db	151	CACATGGAGATGGCGCAAGACCTCGACTATTACCAAAACCATGTCTGGGTGCAAGGAGCA	210
Qy	196	AspPheLysAspLeuGluThrGlnProHisMetClnGluPheArgLeuGlyValArgLeu	215
Db	211	GACTTCAGGATCTTCGAGCCCAAGCCCATATGATGATGTTCTGGCTGGGGGTACCACTC	270
Qy	216	TyrSerGluGluGlnProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTyr	235
Db	271	TACTCAGAGAGAAGCCACAGGAAGCTGTGCCCCACCTGGAGGGCGGCACTGCAAGAGTAC	330
Qy	236	PheValAlaTyrGluGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyr	255
Db	331	TTTGTGCCCATGAGGAGTGCCTGTCCGTCCCTCTGCGAAGGGCCCTATGACTACGACGGCTAC	390
Qy	256	AsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnVal	275
Db	391	AACCTACCTAGCTACAGCGCTGACCTCTTCCAGGCCATCACAGATCATATTAGTCCAGGTC	450
Qy	276	LeuAsnCysLysGlnAsnCysValThrGluLeuAlaSerHisProSerArgGluLysPro	295
Db	451	CTCAACTGTAAAGCAGAACTGTGTCAAGAGCTGGCTTCCCACCACTAGCGGAAAAAGCCC	510
Qy	296	PheGluAspPheLeuProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGly	315
Db	511	TTTGAAAGACTTCCTCCCTTCCACTATAATTAACCTACAGTTTGCTACTACAACATTGGG	570
Qy	316	AsnTyrThrGlnAlaGlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGlu	335
Db	571	AACTATACAAAGCTATTGAATGTGCCAAGACCTACCTCTCTTCTTCCCAATGATGAG	630
Qy	336	ValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSer	355
Db	631	GTGATGACCAGAACTCTGGCTTAATACACAGCCATGCTTGAGAGAAGAGCGGACGCTCC	690
Qy	356	IleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeu	375
Db	691	ATCAGCCCCAGGGAGNN	750
Qy	376	LeuPhePheAlaTyrAspValPheGlyIleProPheValAspProAspSerTrpThrPro	395
Db	751	NN	810
Qy	396	GluGluValIleProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaVal	415
Db	811	GAAGAGGTGATTCCTCAAGAGACTCCAGAGAAACAGAAAGTCGGAACCGCAACCCGAGTC	870
Qy	416	ArgIleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLys	435
Db	871	CGCATCTCCAGGAGATTGGCAACCTTTGAAGAGATCGAGACCTCTCGTGGAGGAGAAG	930
Qy	436	ThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGlu	455
Db	931	ACCAAGAGATCTCTTGATGTGAGCCGACTGACGGGGAAAGTGGTCCCCCTGCTGTATGAA	990
Qy	456	GlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAsp	475
Db	991	GGCATAGTCTCACCATGAATCCAAAGCTCTTGAATGGCTCCCAAGCGGGTGGTATGGAT	1050
Qy	476	GlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThr	495
Db	1051	GGTGTGATCTCTGATGATGAGTGCACAGAGCTGACCAATGGCGGCAACT	1110



```

2 (bases 1 to 1746)
REFERENCE
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,
Ferriter,S., Wang,G., Zheng,X.H., White,T.J., Sniinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source Location/Qualifiers
1..1746
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>1746
/gene="LEPRE1"
/locus_tag="HCM4108"
ORIGIN
US-10-045-815-4 (1-736) x AY411091 (1-1746)
Alignment Scores:
Pred. No.: 2,29e-168 Length: 1746
Score: 1685.00 Matches: 356
Percent Similarity: 61.45% Conservative: 1
Best Local Similarity: 61.27% Mismatches: 224
Query Match: 43.54% Indels: 0
DB: 29 Gaps: 0
156 IleAsnLysLeuGluLysAlaValAlaAlaHisThrPheValGlyAsnProGlu 175
1 ATCAACAAGTTGGAGAAAGCTGTGTGTCAGACACACACCTTCTTCGTGGCAATCCTGAG 60
176 HisMetGluMetGlnGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAla 195
61 CACATGGAATGGCGAGAACCTTAGACTATTACCAACCATGTCTGGAGCGAGGAGGCC 120
196 AspPheLysAspLeuGluThrGlnProHisMetGlnGluPheArgLeuValArgLeu 215
121 GACTTCAAGGATCTTGAGACTCAACCCCATATGCAAGATTTTCGACTGGAGTGGACTC 180
216 TyrSerGluGlnProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyr 235
181 TACTCNAGAAACAGCCACAGAGAGTGTGCCCTCTGCGAAGGCCCTATGACTGANGTGGCTAC 240
236 PheValAlaTyrGluGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyr 255
241 TTGTGGCCTATGAGGAGTGGCTGCCCTCTGCGAAGGCCCTATGACTGANGTGGCTAC 300
256 AsnTyrLeuGluTyrAsnAlaAspLeuPheGlnAlaAlaLeuAspHisTyrIleGlnVal 275
301 AACTACCTTGAGTACACAGCTGACCTCTTCCAGGCCCATCACAGCCCATCACAGCCCATC 360
276 LeuAsnCysLysGlnAsnCysValThrGluLeuAlaSerHisProSerArgGluLysPro 295
361 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 420
296 PheGluAspPheLeuProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGly 315
421 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 480
316 AsnTyrThrGlnAlaGlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGlu 335
481 AATTANNNNVAGCTGTGATGTGCGCAANNCTATCTCTTCNTCCCAATGACGAG 540
336 ValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGluHisThrArgSer 355
541 GTGATGAACCAAAATTTGNNCTATTATGACGATATGCTTGGAGAGAAACACACCAAGTCC 600
356 IleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeu 375
601 ATCGGCCCCCNCNNCNCNCAAGAGTCTCTNNNAGGCGAGTGAANNAGCCCAAGGATGAG 1740

```

```

376 LeuPhePheAlaTyrAspValPheGlyIleProPheValAspProAspSerTyrThrPro 395
661 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 720
396 GluGluValIleProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaVal 415
721 GAAGAGTGTATCCAGAGATTGNNAGAGAAACAGAAANNNNNNNNNNNNNNNNNNNNNNNN 780
416 ArgIleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLys 435
781 NNNNNNNCCAGAGATTGGACCTTATGAAGAAANNNGANNCCCTTGTGGAAGAGAAAN 840
436 ThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGlu 455
841 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 900
456 GlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAsp 475
901 NNNNNNNNNCTCACCATGAACCTCAATGCTTCCAGCGGGTGGTGGATGCAC 960
476 GlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThr 495
961 GGCGTAATCTCTGACCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1020
496 SerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGly 515
1021 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1080
516 ValThrValPheLysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAla 535
1081 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1140
536 HisLeuTyrTyrAsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeu 555
1141 CACTGTACTACACGTGACGAGAGGTGCGGCGCATCATGAGTCTCTACTTCCGCGCTG 1200
556 AspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluVal 575
1201 GATACGCCCTCTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1260
576 GlnAlaGluArgLysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAla 595
1261 CAGCAGAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1320
596 GluThrLeuValCysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIle 615
1321 GAGACCTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1380
616 LeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLys 635
1381 CTTNNCTTAATGGGACTTCGATGGCGAAACTTTTATTTCATCTGAACCTGGATGGCAAG 1440
636 ThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThr 655
1441 ACCGTGACGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1500
656 GluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyr 675
1501 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1560
676 PheThrLeuAspProArgHisSerGluArgAspValGlnAlaAspAspLeuValLys 695
1561 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1620
696 MetLeuPheSerProGluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGln 715
1621 ATGCTCTTCAGCCAGAGAGATGAGCTCTCCAGAGCAGCCCTGGATGCGCCAGCAG 1680
716 GlyProProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGlu 735
1681 GGCCCCCCCCNNCNCNCAAGAGTCTCTNNNAGGCGAGTGAANNAGCCCAAGGATGAG 1740

```



```

QY      736 Leu 736
Db      1741 CTA 1743

RESULT 7
CA488534
LOCUS   AGENCOURT_10809126 MAPcL Homo sapiens cDNA clone IMAGE:6720300 5',
DEFINITION mRNA sequence.
ACCESSION CA488534
VERSION   CA488534.1 GI:24950669
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 999)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
          Email: cgabs@mail.nih.gov
          Tissue Procurement: Kristi A. Eglund, Ira Pastan
          CDNA Library Preparation: Invitrogen Corp
          DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
          Clone Distribution: AGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: L14M14279 row: j column: 12
          High quality sequence stop: 734.
          Location/Qualifiers
            1..999
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:6720300"
              /cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
              hERT-HMEL, LNCaP"
              /lab_host="EMDH108"
              /clone_lib="MAPcL"
              /notes="Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;
              Subtracted with brain, liver, kidney and muscle.
              Directionally cloned. Priming method: oligo-dT. Average
              insert size: 1800 bp. Library amplification: 26,000 fold.
              Kristi A. Eglund, James J. Vincent, Robert Strausberg,
              Bungkok Lee & Ira Pastan: Discovery of new breast
              cancer genes encoding membrane and secreted proteins.
              Manuscript submitted."

FEATURES
source
Alignment Scores:
Pred. No.: 3,77e-162 Length: 999
Score: 1623.00 Matches: 322
Percent Similarity: 97.31% Conservative: 4
Best Local Similarity: 96.12% Mismatches: 6
Query Match: 41.94% Indels: 4
DB: 14 Gaps: 0

US-10-045-815-4 (1-736) x CA488534 (1-999)

QY      277 AsnCysLysGlnAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPhe 296
Db      2 AACTGTAAGCAGAACTGTGTCCAGGAGTTGCTTCCACCCCAAGTCGAGAGAGCCCTTT 61
QY      297 GluAspPheLeuProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsn 316
Db      62 GAAGACTTCTCCCATCGCATTAATATATCTCAGTTTGCCTTCTTCTTCTTCTTCTTCTTCTT 121
QY      317 TyrThrGlnAlaGlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluVal 336
Db      122 TATACACAGGCTGTGAATGTGCCAAGACCTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 181

```

```

QY      337 MetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerile 356
Db      182 ATGAACCAAAATTTGGCTATTATGAGTATGCTTGGAGAGAACACACAGATCCATC 241
QY      357 GlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeu 376
Db      242 GGCCCCCGTGAAGTGGCAAGAGTACCGACAGCAAGCCCTACTGGAAAAAGAACTGCTT 301
QY      377 PhePheAlaTyrAspValPheGlyTyrProPheValAspProAspSerTyrThrProGlu 396
Db      302 TTCTTCGCTTATGATGTTTGGAAATTCCTTTGGATCCGGATTCATGGACTCCAGAA 361
QY      397 GluValIleProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArg 416
Db      362 GAAGTGAATTCCTCAAGAGATTCCAAGAGAAAAGAGAGTTCAGAACCGGAAACAGCGGTACGC 421
QY      417 IleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThr 436
Db      422 ATCTCCAGGAGATTGGAACTTATGAGGAAATCGAGACCTTTGTGGAGAGAGAACACC 481
QY      437 LysGluSerLeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGly 456
Db      482 AAGGAGTCACTGGATGTGAGCAGACTGACCCCGGAAGTGCCCTCTGCTATGAAGGC 541
QY      457 IleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAspGly 476
Db      542 ATCAGTCTCACCATGAATCTCCAACTCTGTAATGTTCCCAAGCGGGTGTGATGGACGCG 601
QY      477 ValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSer 496
Db      602 GTAATCTCTGACCACGAGTGTGAGGAGCTGCAGAGACTGACCAATGTGGCAGCAACTCA 661
QY      497 GlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyVal 516
Db      662 GGAGATGGCTACCGGGGTGACAGCTCCCACTACTCCCAATGAAAGTTCTTATGGTGTCT 721
QY      517 ThrValPheLysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHis 536
Db      722 ACTGTCTCAAGCC-CtCAAGCTGGGCAAGAGCAAGTTCCTCTCTCAGAGTGGCCAC 780
QY      537 LeuTyrTyrAsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAsp 556
Db      781 CTGTACTACAAAGTGCAGCAGAGAGTGGCGGCATCATGGAGTCTCTACTTCGGCTGGAT 840
QY      557 ThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluVal-Gl 576
Db      841 ACGCCCTCTACTTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
QY      576 nAlaGluArgLysAspAspSerHisProValHisValAspAsnCysIleLeuAsn-Alag 596
Db      901 GGCANANAGGAAGGATGATAGTCTTCCAGTCCAGTGGACACTGCATCTCTGAATGGCCA 960
QY      596 luThrLeuValCysVal-LysGluProProAlaTyrThr 608
Db      961 AGACCTCTCTGTGTCAAAAGAGGCGCCAGCTACACN 999

RESULT 8
LOCUS   BUI48566
DEFINITION BUI48566 899 bp mRNA linear EST 03-SEP-2002
          AGENCOURT_8675673 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6380366
          5', mRNA sequence.
ACCESSION BUI48566
VERSION   BUI48566.1 GI:22662098
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 899)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.

```

Email: cgapbs-remail.nih.gov  
Tissue Procurement: DCTD/DTP  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CM2568 row: n column: 15  
High quality sequence stop: 724.

FEATURES  
source  
1..899  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6380366"  
/tissue\_type="carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_40"  
/note="Organ: prostate; Vector: pOTB7; Site: 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dr priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

ORIGIN  
Alignment Scores:  
Pred. No.: 1.69e-152 Length: 899  
Score: 1532.00 Matches: 293  
Percent Similarity: 98.33% Conservative: 1  
Best Local Similarity: 97.99% Mismatches: 5  
Query Match: 39.59% Indels: 1  
DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x BUI48566 (1-899)

Qy 143 ArgSerProTyrAsnTyrLeuGlnValAlaTyrPhelysileAsnLysLeuGluYsAla 162  
Db 3 CGGAGCCCTTACAACTACCTGCGAGTCCCTTCTTCAAGATCAACAGTTGGAGAAGCT 62

Qy 163 ValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGlnGlnAsn 182  
Db 63 GTTCTGTCAGCACACACTTCTTCTGCGGCACTCTGAGCACATGGAATGACGAGAAC 122

Qy 183 LeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeuGluThr 202  
Db 123 CTAGACTATTACCAAAACCAATGTCTGGAGTGAAGAGCCGACATCAAGGATCTTGAGCT 182

Qy 203 GlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGlnProGln 222  
Db 183 CAACCCCATATGCAGAAATTCGACTGGAGTGGACTCTACTCAGAGGACAGCCACAG 242

Qy 223 GluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGluGluCys 242  
Db 243 GAAGCTGTGCCCCACCTAGAGGGCGGCTCAAGAAATCTTGTGGCCCTATGAGGAGTGC 302

Qy 243 ArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAla 262  
Db 303 CGTGCCCTCTCGAGAGGCCCTATGACTAGTGGCTAGCACTACCTTGAGTACACGCT 362

Qy 263 AspLeuPheGlnAlaLalleThrAspHisTyrIleGlnValLeuAsnCysLysGlnAsnCys 282  
Db 363 GACCTCTCCAGGCCATCACAGATCATATCATCCAGTCTCTCACTCACTGTAAGCAACTGT 422

Qy 283 ValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeuProSer 302  
Db 423 GTACGGAGCTTGCTTCCACCCCAAGTCGAGAGAGCCCTTGAAGACTTCTCCCATCG 482

Qy 303 HisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAlaGlyGlu 322

Db 483 CATTATAATATCTCCAGTTTGCCTACTATTAACATTTGGGAATTTATACACAGCTGTTGAA 542

Qy 323 CysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsnLeuAla 342

Db 543 TGTGCCAAGACCTATCTTCTTCTTCCCAATGACGAGGTGATGAATCAAAATTTGGCC 602

Qy 343 TyrTyrAlaAlaMetLeuGlyGluHisThrArgSerIleGlyProArgGluSerAla 362

Db 603 TATTATGAGCTATGCTTGGAGAAACACACAGATCCATCGCCCGGCGAGAGTGCC 662

Qy 363 LysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhePheAlaTyrAspVal 382

Db 663 AAGAGTACCCAGCAGGAGCTACTGGAAAAGAACTGCTTTCTTCGTATATGTT 722

Qy 383 PheGlyIleProPheValAspProAspSerTrpThrProGluGluValIleProLysArg 402

Db 723 TTTGGAATTCCTTTGTGATCCGATTCATGGACTCCAGAAGAAGTGAATTCCTCAAGAGA 782

Qy 403 LeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGluIleGly 422

Db 783 TTGCAAGAGAAACAGAGTCAAGCGGAAACAGCCGTACGCACTCCAGAGGATGGG 842

Qy 423 AsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeuAsp 441

Db 843 AACTT-ATGAAGGAATCGAGACCTTGTGGAAGAGAGACAAAGAGTCACTGGAT 898

RESULT 9  
BX364036 1101 bp mRNA linear EST 05-MAY-2003  
BX364036 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
Homo sapiens cDNA clone CS0DL008YM13 5-PRIME, mRNA sequence.

ACCESSION BX364036  
VERSION BX364036.1 GI:30370766  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 476.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DL008AG07QPI&cluster=476.r. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DL008AG07QPI.

FEATURES  
source  
1..1101  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DL008YM13"  
/cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN  
Alignment Scores:  
Pred. No.: 4.74e-150 Length: 1101  
Score: 1510.50 Matches: 300

[illegible]



QY 578 GluA-GlyAspSerHisProValHisValAspAsnCysIleLeu-AsnAlaGluTh 597  
 Db 828 AGAAGAA-GATCATAGTCATCCAGTCCACGTGGACACTGCATCTGAAATGCCAAGAC 886  
 QY 597 rLeuValCysValLysGluProAlaTyrThr-PheArgAspTyrSerAlaIleLeuT 617  
 Db 887 CTTGCTGTGTGTAAGAGCCCGAGCT---ACACTTCCGCACATACAGGGCATCTTTA 943  
 QY 617 rLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrV 637  
 Db 944 CTTAAAGGGGGGACTTCGATGCGGGAACCTTTATTTCCT-GAATCGATGCCAAACCGG 1002  
 QY 637 alThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPheSer---SerGlyThrG 656  
 Db 1003 TGACGGCAAGGGGCCACCCCGCCAGCGGGGAAAAACCGGGGGAATCTCTTTCCGGGCTGG 1062  
 QY 656 luAsnProHis 659  
 Db 1063 AAACCCCCAC 1073

RESULT 12  
 BX334235 1201 bp mRNA linear EST 01-MAY-2003  
 LOCUS BX334235 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
 DEFINITION clone CSODI001YG09 5-PRIME, mRNA sequence.  
 ACCESSION BX334235  
 VERSION BX334235.1 GI:30310318  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by life technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 476.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CSODI001AD05QPI&cluster=476.r. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CSODI001AD05QPI.  
 FEATURES  
 source  
 1..1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CSODI001YG09"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 4,96e-146 Length: 1201  
 Score: 1473.50 Matches: 292  
 Percent Similarity: 96.69% Conservative: 0  
 Best local Similarity: 96.69% Mismatches: 9  
 Query Match: 38.07% Indels: 4  
 DB: 13 Gaps: 0

US-10-045-815-4 (1-736) x BX334235 (1-1201)

QY 5 AlaLeuLysLeuLeuThrLeuLeuAlaValAlaAlaAlaLaserGlnAlaGluVal 24  
 Db 109 GCGTTTGAAGCTGTGACACACACTGCTGGCTGTGGTGGCGCTGCCCTCCCAAGCGAGGTC 168  
 QY 25 GluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAlaGluGlyThrAla 44  
 Db 169 GAGTCCGAGGAGGATGGGCATGCTGACGCTGATCTGCTTTCGCCGAGGGAGCGCA 228  
 QY 45 AlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArgAlaLeuArgSer 64  
 Db 229 GCCTACGCGCGGGGAGTGCCTCCGGGTGGTCTGAGCATGGAAACGGGGCGCTCGCTGCC 288  
 QY 65 ArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAlaAlaAspPhePro 84  
 Db 289 CGGCGAGCCCTCCGCGCCCTTCGCTGCGTCCGCAACCCAGTGTGCGCGGACTTCCTCCG 348  
 QY 85 TrpGluLeuAspProAspTyrPheSerProAlaGlnAlaSerGlyAlaGlyAlaLeu 104  
 Db 349 TGGGAGCTGGACCCCGACTGTGTCCCGAGCCCGCCAGGCTCGGGCGCCGCCCTG 408  
 QY 105 ArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLeuArgCysLeu 124  
 Db 409 CGGACCTGAGCTTCTTCGGGGGCTTCTGCTGCGCTGCTGCTGCTGCGCGCTGCTGCTC 468  
 QY 125 GlyProProAlaAlaHisSerLeuSerGluLeuMetGluLeuGluPheArgSer 144  
 Db 469 GGGCGCGCGCGCCCACTGCTCAGGAGAGATGGAGCTGGAGTTCCTCCGACGGGAGC 528  
 QY 145 ProTyrAsnTyrLeuGlnValAlaTyrPheLysIleLeuLysLeuGluLysAlaValAla 164  
 Db 529 CCTTCAACTACTGCTGAGTGCCTTCTCAAGATCAACAAGTTGGAGAAAGCTTGTCT 588  
 QY 165 AlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGlnGlnAsnLeuAsp 184  
 Db 589 GCAGCACACACCTTCTTCGTGGCAATCTGAGCACATGGAATGCAGCAGAACCTAGAC 648  
 QY 185 TyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeuGluThrGlnPro 204  
 Db 649 TATTACCAACCATCTCTGGAGTGAAGAGCGCGACTTCAAGGATCTTGAGACTCAACCC 708  
 QY 205 HisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGlnProGlnGluAla 224  
 Db 709 CATATGCAAGAAATTCGACTGGAGTGGAGTCTACTCAGAGGAACAGCCACAGGAAGCT 768  
 QY 225 ValProHisLeuGluAlaAlaLeuGlnLutyrPheValAlaTyrGluGluCysArgAla 244  
 Db 769 GTGCCCCACCTAGAGCGCGCTGCAAGAATACTTTGTGGCTATGAGGAGTGGCGTGC 828  
 QY 245 LeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyrAsnAlaAspLeu 264  
 Db 829 CTCCTGGAAGGGCCCTATGAMTANGATGGCTACAACCTACCTTGACTACAACGCTGACCTC 888  
 QY 265 PheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln-AsnCysValTh 284  
 Db 889 TTCCAGGCCCATCACAGATCAITACATCCAGGTCCTCAACTGTAARCAAGAACTGTGTAC 948  
 QY 284 rGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeuProSerHisTyr 304  
 Db 949 GGAGCTTGTCTTCAA-CCAAAGTCGAGAGNASCCTTT-GAAGACTTCTT-CCATYGSATTA 1005  
 QY 304 rAsn 305  
 Db 1006 TAAT 1009

RESULT 13  
 BX442345 1201 bp mRNA linear EST 15-MAY-2003  
 LOCUS BX442345 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone  
 DEFINITION CSODF029YB14 5-PRIME, mRNA sequence.  
 ACCESSION BX442345  
 VERSION BX442345.1 GI:30776018  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)

**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 1201)  
**AUTHORS** Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
**TITLE** Full-length cDNA libraries and normalization  
**JOURNAL** Unpublished (2001)  
**COMMENT** Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 476.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0DF029DA07QPI&cluster=476.r. Contact :  
 Feng Liang Email : fliang@lifetech.com URL : Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0DF029DA07QPI.  
**FEATURES**  
 Location/Qualifiers  
 1..1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DF029YB14"  
 /tissue\_type="FETAL BRAIN"  
 /dev\_stage="fetal"  
 /clone\_lib="Homo sapiens FETAL BRAIN"  
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
 was primed with a NotI-oligo(dT) primer. Five prime end  
 enriched, double-strand cDNA was digested with NotI and  
 cloned into the NotI and EcoRV sites of the pCMVSPORT 6  
 vector. Library was not normalized."  
**ORIGIN**  
 Alignment Scores:  
 Pred. No.: 1,568-144 Length: 1201  
 Score: 1459.50 Matches: 299  
 Percent Similarity: 91.84% Conservative: 5  
 Best Local Similarity: 90.33% Mismatches: 22  
 Query Match: 37.71% Indels: 8  
 DB: 13 Gaps: 2  
 US-10-045-815-4 (1-736) x BX442345 (1-1201)  
 QY 1 MetAlaValArgAlaLeuLysLeuLeuThrLeuLeuAlaValValAlaAlaSer 20  
 Db 89 ATGCGGTACGCGGTTGAAGCTGCTGACCACACTGCTGCTGCGCGCTGCTCC 148  
 QY 21 GlnAlaGluValGluSerGluAlaGlyTTPGlyMetValThrProAspLeuLeuPheAla 40  
 Db 149 CAACCCGAGGTGAGTCCGAGGAGGATGGGATGATGAGCTGATCTGCTTCGCGC 208  
 QY 41 GluGlyThrAlaAlaValAlaArgGlyAspTTPProGlyValValLeuSerMetGluArg 60  
 Db 209 GAGGGACCCGACGCTACGCGCGCGGGAGCTGCGCGGGTGGTCTGAGCATGGAACGG 268  
 QY 61 AlaLeuArgSerArg-AlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAl 80  
 Db 269 GCCTGGCTCCCGGGTACCCCTCCGCGCCCTTCGCTGCGTGCAGCCAGGTGTC 328  
 QY 80 aAlaAspPheProTTPGluLeuAspProAspTTPSerProSerProAlaGlnAlaSerG1 100  
 Db 329 CGCCGACTTCCCTGGGAGTGGACCCCGACTGGTCCCGCAGCCGCGCGCTCGGG 388  
 QY 100 ValAlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLe 120  
 Db 389 CGCGCGCGCCCTCGCGACCTGAGCTTCTTCGGGGGCGCTTCGCGTGGCTGCGCT 448  
 QY 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140  
 Db 449 GCGCGCGTGCCTCGGCGCGCGCGCCCGCCACTCGCTCAACGAAATAATFAAGCTGGAGT 508

QY 140 eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuG1 160  
 Db 509 CCCAAGCGGAC-CCCTCAACTACMTGAGGTGCGCTACTTCAAGATCAACAAATTGGA 567  
 QY 160 uLysAlaValAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetG1 180  
 Db 568 GAAAGCTGTTGTCGAGCAGACACACCTTCTTCTGGCAATCTCTGAGCATGGAATGCA 627  
 QY 180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
 Db 628 GCAGAACCTAGACTATTACCAACCATGCTCTGAGTGAAGGAGGCGGACTTCAAGGATCT 687  
 QY 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValAlaGlyLeuTyrSerGluGluG1 220  
 Db 688 TGAGACTCAACCCCATATGCAAGAAATTCGACTGGAGTGGCGACTCTACTCAGAGGAACA 747  
 QY 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyr-G 240  
 Db 748 GCACAGGAGGCTGTGCCCCACTAGAGCGCGCTGCAAGAAATACTTTGTGGCCTATGG 807  
 QY 240 LuGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluT 260  
 Db 808 AGAGTGCGCGTCCCTCTGCGAAGGCGCTATGACTAGTGGCTACAACTACCTTGAGT 867  
 QY 260 YrAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysG 280  
 Db 868 ACAACGCTGACCTCTCCAGGCGCATCACAGATCATTCATCCAGGCTCCTCAACTGTAARC 927  
 QY 280 InAsn-CysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPhe 299  
 Db 928 AGAACCTGTGTACCGAGCTTCTKCCCAACCAAGTCGAGAGAGCCSTTG---ARACTT 984  
 QY 300 LeuProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGln 319  
 Db 985 CTKCCATCGCATATAW-TATCTGAGTT-CCYWTATWATTTGGGATGGAARAACA--- 1039  
 QY 320 AlaGlyGluCysAlaLysThrTyrLeu 328  
 Db 1040 GGYGTTGAATGKCAARMCTATYTCY 1066  
**RESULT 14**  
**AL521774**  
**LOCUS** AL521774 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens  
**DEFINITION** cDNA clone CS0DB003YL11 5-PRIME, mRNA sequence.  
**ACCESSION** AL521774  
**VERSION** AL521774.2 GI:31040051  
**KEYWORDS** EST.  
**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 926)  
**AUTHORS** Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
**TITLE** Full-length cDNA libraries and normalization  
**JOURNAL** Unpublished (2001)  
**COMMENT** On Feb 13, 2001 this sequence version replaced gi:12785267.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 476.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0DB003CF06QPI&cluster=476.r. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0DB003CF06QPI.  
**FEATURES**  
 Location/Qualifiers  
 1..926  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"



```
Db      180 ACCAGGAGTCACTGGATGTGAGCAGACTGACCCCGGAGAGTGGCCCCCTGCTGTATGAA 239
Qy      456 GlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyrGlnArgValValMetAsp 475
Db      240 GGCATCAGTCTCACCATACTCCAACTCCCTGAATGGTCCAGCGGGTGGTGTATGGAC 299
Qy      476 GlyValIleSerAspHisGlnCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThr 495
Db      300 GCGGTAACTCTGACACACAGAGTTCAGGAGCTGCAGAGACTGACCAATGTGGCAGCAACC 359
Qy      496 SerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGly 515
Db      360 TCAGGAGATGGCTACCGGGGTCCAGACCTCCCCACATACTCCCAATGAAAAGTTCTATGGT 419
Qy      516 ValThrValPheLysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAla 535
Db      420 GTCACCTGCTTCAAAGCCCTCAAGCTGGGGCAAGAAGCAAAAGTTCCCTCTGCAGAGTGCC 479
Qy      536 HisLeuTyrTyrAsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeu 555
Db      480 CACCTGTACTACACGTGACGGAGAGAGGTGGCGGCATCATGGAGTCTTACTTCCGCTG 539
Qy      556 AspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluVal 575
Db      540 GATAGCCCCCTCTACTTTCTCTACTCTCTCATCTGGTGTCCGCACTGCATCGAAGAGGTC 599
Qy      576 GlnAlaGluArgLysAspSerHisProValHisValAspAsnCysIleLeuAsnAla 595
Db      600 CAGGACAGAGAGAGAGGATGATGATCCAGTCCAGTGGACAACTGCATCCTGAATGCC 659
Qy      596 GluThrLeuValCysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIle 615
Db      660 GAGACCTCTGCTGTGTCAAAGAGCCGCCAGCCTACACCTTCGCGCACTACAGCGCCATC 719
Qy      616 LeuTyrLeuAsnGlyAspPheAspGlyGlyAsn-PheTyrPheThrGluLeuAspAlaLys 635
Db      720 CTTTACCTAAATGGGACCTTCGATGGCGGAACTTTTATTTCACCTGAATGGATGCCAA 779
Qy      635 sThrValThrAlaGluValGlnProGlnCysGly-ArgAlaValAlGlyPheSerSer-Gly 654
Db      780 GACCGTGACGGCAGAGGTGCAGGCTCAGTGTGGAAGAGCGGTGGGATTTCTTCCAGGC 839
Qy      655 ThrGlu-AsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLe 674
Db      840 ACTGAAACCCCATGGAGTGAAGGCTGTACCCAGGGGCAAGGCTGTGC---ATTGCGCT 896
Qy      674 uTrpPheThrLeuAsp 679
Db      897 GTGGTTCCACCTCGAA 912
```

Search completed: July 18, 2004, 17:57:01  
Job time : 6590 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: July 18, 2004, 16:08:41 ; Search time 10733 Seconds  
(without alignments)

2972.186 Million cell updates/sec

US-10-045-815-4

Title:

Perfect score: 736

Sequence: 1 MAVRALKLLTLLAVVAAS.....PPEPAQESLSGESKPKDEL 736

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6934743

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2 1/USPTO.spool/p/US10045815/runat 14072004 123119 17240/app\_query.fasta\_1.903  
-DB=GenEmbl -OFMT=fastap -SUFFIX=oli.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10045815@cgn 1 17406 @runat 14072004 123119 17240 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	736	100.0	2600	9 AF097432	AF097432 Homo sapi
2	433	58.8	2563	6 AX136191	AX136191 Sequence
3	433	58.8	2563	6 BD123548	BD123548 Secretary
4	433	58.8	2563	9 AK075418	AK075418 Homo sapi
5	433	58.8	2583	6 AX704765	AX704765 Sequence
6	433	58.8	2585	6 AX877288	AX877288 Sequence
7	433	58.8	2585	6 BD156581	BD156581 Primer fo
8	433	58.8	2585	9 AK027697	AK027697 Homo sapi
9	407	55.3	2524	6 AX877139	AX877139 Sequence
10	407	55.3	2524	6 BD156505	BD156505 Primer fo
11	407	55.3	2524	9 AK027680	AK027680 Homo sapi
12	360	48.9	2829	9 AF097431	AF097431 Homo sapi
13	301	40.9	1562	9 BC015309	BC015309 Homo sapi
14	295	40.1	2993	6 AX880254	AX880254 Sequence
15	295	40.1	2993	6 BD158280	BD158280 Primer fo
16	285	40.1	2993	9 AK027648	AK027648 Homo sapi
17	289	39.3	1173	9 BT007039	BT007039 Homo sapi
18	289	39.3	1173	12 BT007768	BT007768 Synthetic
19	204	27.7	1526	9 AK025841	AK025841 Homo sapi
20	179	24.3	550	6 AX677402	AX677402 Sequence
21	103	14.0	682	6 BD020571	BD020571 Novel gen
22	103	14.0	682	6 BD100509	BD100509 Novel gen
23	101	13.7	652	6 AX136510	AX136510 Sequence
24	101	13.7	652	6 BD123750	BD123750 Secretary
25	101	13.7	738	6 AX868542	AX868542 Sequence
26	101	13.7	738	6 BD148604	BD148604 Primer fo
27	101	13.7	774	6 AX868815	AX868815 Sequence
28	101	13.7	774	6 BD148877	BD148877 Primer fo
29	101	13.7	2524	10 AF087433	AF087433 Rattus no
30	101	13.7	2569	10 BC024047	BC024047 Mus muscu
31	101	13.7	161835	9 AC098484	AC098484 Homo sapi
32	91	12.4	2297	10 AF165164	AF165164 Mus muscu
33	91	12.4	2391	10 AF165163	AF165163 Mus muscu
34	80	10.9	703	6 AX868722	AX868722 Sequence
35	80	10.9	703	6 BD148784	BD148784 Primer fo
36	73	9.9	711	6 BD020572	BD020572 Novel gen
37	73	9.9	711	6 BD100510	BD100510 Novel gen
38	58	7.9	569	6 AX873839	AX873839 Sequence
39	58	7.9	569	6 BD153901	BD153901 Primer fo
40	47	6.4	193219	10 AL606975	AL606975 Mouse DNA
41	47	6.4	194020	2 AC101988	AC101988 Mus muscu
42	47	6.4	225308	2 AC098918	AC098918 Rattus no
43	42	5.7	554	6 AX873755	AX873755 Sequence
44	42	5.7	554	6 BD153817	BD153817 Primer fo
45	37	5.0	162	6 AX916981	AX916981 Sequence

# ALIGNMENTS

RESULT 1

AF097432 2600 bp mRNA linear PRI 22-NOV-2000  
 LOCUS Homo sapiens GROSI-L protein mRNA, complete cds.  
 DEFINITION AF097432  
 ACCESSION AF097432  
 VERSION AF097432.1 GI:11127637  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 2600)  
 AUTHORS Kaul, S.C., Sugihara, T., Yoshida, A., Nomura, H. and Wadhwa, R.  
 TITLE GROSI, a potential growth suppressor on chromosome 1: its identity  
 to basement membrane-associated proteoglycan, leprecan  
 JOURNAL Oncogene 19 (32), 3576-3583 (2000)  
 MEDLINE 20406537  
 PUBMED 10951563  
 REFERENCE 2 (bases 1 to 2600)  
 AUTHORS Wadhwa, R., Sugihara, T. and Kaul, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-OCT-1998) Group 4, Chugai Research Institute of  
 Molecular medicine, 153-2 Nagai, Niihari, Ibaraki 300-41, Japan  
 FEATURES  
 source  
 1..2600  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /tissue\_type="testis"  
 52..2262  
 /note="member of protein B family"  
 /codon\_start=1  
 /product="GROSI-L protein"  
 /protein\_id="RAG31019.1"  
 /db\_xref="GI:11127638"  
 /translation="MAVRLKLLTLAVVAASQAQVESEAGKGMVTPDLLPAEGTA  
 AYARDGFWGLSERRALRLRCRTQCAADFWEIDPDWSPQAQAG  
 ALRDLSPFGGLREACLRCLGPPAAHLSSEMELEFRKRSPTNYLQVAFKINKLE  
 KAVAAHTFVGNPEHMEMQNLIDYQTMGSKVEADPKLETPHMOERFLGVRLYSE  
 EPOEAAPHLEALAEQVEFAVEECRLCEGPDYDGVNLYENADLFOAITDHIQVL  
 NCKQNCVTELASHPRSKPEDELP SHYNYLQFAYNIGNYTAGSCAKTYLLFPND  
 EVMQNLAYAMLESHTSIGPRAKRYRQSLERKLELFFADVFGIPFVDPDS  
 WTPVEIPKRLQKQSERETAVRSOIEIGNKMBIETLVEEKTKESLDVSRLTRGG  
 PLLYEGISLTNMSKLNGYQVYMDGVI SDHECEQLRLTNVAATSDGVRGTSPT  
 PNEKFGYGVTFKALQGEKVPLOSAHLYINVTKEVRIMESYFLDTPLYFSYHL  
 VCRTAIEVQAEKRDSDHPVNDICLNASTLVCKVEPPAYTPRDSAILYLNGDFDG  
 GNFTFLDAKTVTAEVQPOCGRAVGFSGTENPHGVKAVTRGRCRAIALWFTLDPRH  
 SERDEVDADLVKMLFSPPEMDLSQEPDLAQGPPEPAQESLSGESKPKDEL"

Alignment Scores:  
 Pred. No.: 0 Length: 2600  
 Score: 736.00 Matches: 736  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-10-045-815-4 (1-736) x AF097432 (1-2600)

QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
 Db 52 ATGGCGGTACCGCGTTGAAGCTGCTGACACACTGCTGCTGCTGGCGCGCTGCTCC 111  
 QY 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuLeuPheAla 40  
 Db 112 CAAGCCGAGTCCGAGTCCGAGGAGGATGGGCGATGGCGCTGCTGCTCTTCGCC 171  
 QY 41 GluGlyThrAlaAlaValAlaArgGlyValAspTyrProGlyValValLeuSerMetGluArg 60  
 Db 172 GAGGGGACCGGAGCTTACGCGCGGGGAGCTGGCGCGGGTGGTCTTGACATGGAGACGG 231  
 QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80

232 GGCGTCCGTCCTCCGGGAGCCCTCCGGCGCTTCGCCCTGCGCTGCCACCCAGTGTGCC 291  
 QY 81 AlaAspPheProTyrGluLeuAspProAspTyrPserProAspPheAlaGlnAlaSerGly 100  
 Db 292 GCGGACTTCCCGTGGAGCTGGACCCCGACTGGTCCCCCAGCCCGCCGCGCTCGGGC 351  
 QY 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120  
 Db 352 GCGGCGCCCTCCGGGAGCTGGACCCCGACTGGTCCCCCAGCCCGCCGCGCTCGGGC 411  
 QY 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuPhe 140  
 Db 412 CGCCGCTGCTCGGCGCGCGCGCCGCCACCTCTCTCGGGGCAATCTCTGACGAGAGATGGAGTTC 471  
 QY 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160  
 Db 472 CGCAAGCGGAGCCCTACAACTACCTGAGTGGCTGCTCAAGATCAACAAGTTGGAG 531  
 QY 161 LysAlaValAlaAlaAlaHisThrPheValGlyAsnProGluHisMetGluMetGln 180  
 Db 532 AAAGCTGTGCTGCGAGCACACACCTCTCTCGGGGCAATCTCTGACGAGATGGAAATGAG 591  
 QY 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200  
 Db 592 CAGAACCTAGACTATTACCAACCATGCTGAGTGAAGAGGCGGCGACTTCAAGGATCTT 651  
 QY 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGln 220  
 Db 652 GAGACTCAACCCCATATCAAGATTTTCGACTGGGAGTGGCGACTCTACTCAGAGGAGACAG 711  
 QY 221 ProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
 Db 712 CCACAGGAGAGCTGTGCCCCACCTAGAGGGCGGCTGCAAGAACTATTGTGCGCTATGAG 771  
 QY 241 GluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
 Db 772 GAGTCCGCTGCCCTCGGAAAGGCGCTTACTGACTACGATGGCTACCACTACTGAGTAC 831  
 QY 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysGln 280  
 Db 832 AACGCTGACCTCTTCAGGCCATCACAGATCATTACATCCAGGTCCTCACTGTAAGCAG 891  
 QY 281 AsnCysValThrGluLeuAlaSerHisProSerArgGlyLysProPheGluAspPheLeu 300  
 Db 892 AACTGTGTCCGAGCTTGTCTCCACCAAGTCGAGAGAGAGCCCTTTGAAGACTTCCTC 951  
 QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
 Db 952 CCATCGCATTAATATCTGAGTTTGCCTACTATAACATTTGGAAATATACACAGCT 1011  
 QY 321 GlyGluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
 Db 1012 GGTGAATGTGCCAAGACCTATCTTCTTCTTCCCAATGACGAGGTGATGAACCAAAAT 1071  
 QY 341 LeuAlaTyrTyrAlaAlaMetLeuGlyGluHisThrArgSerIleGlyProArgGlu 360  
 Db 1072 TTGGCTATTATGACGCTATGCTTGGAGAGAAACACACAGATCCATCGGCCCGCGTGG 1131  
 QY 361 SerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTyr 380  
 Db 1132 AGTGCAAGGAGTACCGACAGGAGGAGCTACTGGAAAAAGAACTGCTTTCTTCGCTTAT 1191  
 QY 381 AspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIlePro 400  
 Db 1192 GATGTTTTTGGAAATTCCTTTTGGATCCGGAATTCATGGACTCCAGAGAGAGTATCCC 1251  
 QY 401 LysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnGlu 420  
 Db 1252 AAGAGTTTGCAGAGAGAAACAGAGTCAAGACGGGAAACAGCGTACGCACTCCCGAGGAG 1311  
 QY 421 IleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerLeu 440



QY 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyValAsnProGluHisMetGluMetG1 180  
DB 510 GAAAGCTTGTCTGAGACACACACCTTCTTCTGGGCAATCTCTGAGCACATGGAAATGCA 569  
QY 180 nGlnAsnLeuAspTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
DB 570 GCAGAACCTAGACTATTACCAAAACCATCTCTGGAGTGAAGGAGGCGGACCTTCAAGGATCT 629  
QY 200 uGlnThrGlnProHisMetGlnGluPheArgGluGlyValArgLeuTyrSerGluGluG1 220  
DB 630 TGAGACTCAACCCCATATGCAAGAAATTTGAGTGGAGTGGACTTACTCAGAGGAACA 689  
QY 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrG1 240  
DB 690 GCCACAGGAAGCTGTGCCCCACCTAGAGGCGCGCTGCAAGAAATCTTTTGTGGCTATGA 749  
QY 240 uGlnCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
DB 750 GGAGTGGCGTCTGCTGCGAAGGCGCTATGACTAGCATGGCTTACCACTACCTTTGAGTA 809  
QY 260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysG1 280  
DB 810 CAACGCTGACCTCTTCCAGGCGCATCACAGATCATTTACATCCAGTCTCAACTGTAGCA 869  
QY 280 nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
DB 870 GAACCTGTGTCCAGGAGCTTGTCTCCACCCCAAGTCGAGAGAGCCCTTTTGAAGACTTCCT 929  
QY 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyValAsnTyrThrGlnAl 320  
DB 930 CCCATCGCATTAATTAATCATCTGAGTTGCTTACTATAACATTGGGAATATATACAGGC 989  
QY 320 aGlyCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
DB 990 TGT-TGAATGTGCCAAGACCTTATCTTCTTCTTCTTCCCAATGACGAGGTGATGAACCAA 1048  
QY 340 snLeuAlaTyrTyrAlaAlaMetLeuGluGluHisThrArgSerIleGlyProArgG 360  
DB 1049 ATTTGGCTTATTATGAGCTATGCTTGGAGAAGAACACACAGATCATCGGCCCCCGTG 1108  
QY 360 luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaT 380  
DB 1109 AGAGTGCCAGGAGTACCGACAGCGAAGCTTACTTGGAAAAGAACTCTTTTCTTCTGCTT 1168  
QY 380 yrAspValPheGlyIleProPheValAspProAspSerThrThrProGluGluValIleP 400  
DB 1169 ATGATGTTTTTGGAAATTCCTTTTGTGATCGGATTCATGACTCCAGAAAGAGTGAATC 1228  
QY 400 roLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
DB 1229 CCAAGAGATTGCAAGAGAAACAGAGTACAGCGGAAACAGCGTACGATCTCCAGG 1288  
QY 420 luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440  
DB 1289 AGATTGGGAACCTTATGAAGAAATCAGACCCCTTGTGGAAGAGAAACCAAGAGTCTAC 1348  
QY 440 euAspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyIleSerLeuT 460  
DB 1349 TGGATGTGAGAGACTGACCCGGGAAGTGGCCCCCTGCTGTATGAGGATCATGCTCTCA 1408  
QY 460 hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer 479  
DB 1409 CCATGAACCTCAAACTCCTGAATGGTTC-CCAGCGGTGGTGTGAGGAGCGGTAATCTCT 1467  
QY 480 AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499  
DB 1468 GACCACGAGTGTGAGGAGCTGCAGAGACTGACCAATGTGCACGACACCTCAGGAGATGGC 1527  
QY 500 TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe 519  
DB 1528 TACCGGGGTGAGACCTCCCACTATCTCCCAATGAAAAGTTCTATGGTGTCACTGTCTTC 1587

QY 520 LysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyr 539  
DB 1588 AAAGCCCTCAAGCTGGGCAAGAGGCAAGTTCCTCTGCAGAGTGGCCACTGTACTTAC 1647  
QY 540 AsnValThrGluLysValArgIleMetGluSerTyrPheArgLeuAspThrProLeu 559  
DB 1648 AACGTGACGAGGAAGTGGCGGCATCATGGAGTCTTCTTCCGCTGGATACGCCCTC 1707  
QY 560 TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGluAlaGluArg 579  
DB 1708 TACTTTTCTTACTCTCACTGCTGCGCACTGCATCGAAGAGTCCAGCAGAGAGG 1767  
QY 580 LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599  
DB 1768 AAGGATGATAGTCAATCCAGTCCAGTGGACAACTCATCTGAATGCCAGACCCCTCGT 1827  
QY 600 CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619  
DB 1828 TGTGTCAAGAGCCCCCAGCCTTACCTTCCGCGACTACAGCCCATCTTTTACCCTAAT 1887  
QY 620 GlyAspPheAspGlyGlyValAsnPheTyrPheThrGluLeuAspAlaLysThrValThra 639  
DB 1888 GGGGACTTCGATGGCGGAAACTTTTATTTCTCACTGAATGGATGCCAGACCGTGACGCA 1947  
QY 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659  
DB 1948 GAGGTGCAGCTCAGTGTGGAAGCCGTGGATTTCTTTCAGGCACCTGAAAACCCCAT 2007  
QY 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAsp 679  
DB 2008 GAGGTGAAGCTGTCAACAGGGGGGAGCGCTGTGCCATCGCCTGTGTTTCACTTGGAC 2067  
QY 680 ProArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMetLeuPheSer 699  
DB 2068 CCTCGACACAGCGAGCGGACAGGCTGTCAGGAGAGTACCTTGGTGAAGATGCTCTTTCAGC 2127  
QY 700 ProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProGlu 719  
DB 2128 CCAGAAGAGATGAGCTCTTCCAGGAGAGCCCTTGTATGCCAGCAGGGCCCCCGGAA 2187  
QY 720 ProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
DB 2188 CCTGCACAAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2238  
RESULT 3  
BD123548  
LOCUS  
DEFINITION  
SECRETORY protein or membrane protein.  
ACCESSION  
BD123548  
VERSION  
BD123548.1 GI:23218493  
KEYWORDS  
JP 2002017376-A/57.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 2563)  
AUTHORS  
Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and  
Hayashi,K.  
TITLE  
Secretory protein or membrane protein  
JOURNAL  
Patent: JP 2002017376-A 57 22-JAN-2002;  
HELIX RESEARCH INSTITUTE  
COMMENT  
OS Homo sapiens (human)  
PN JP 2002017376-A/57  
PD 22-JAN-2002  
PF 07-JUL-2000 JP 2000253173  
PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,YURI KAWAI,TOMOYASU  
PI SUGIYAMA,  
PI KOJI HAYASHI  
PC  
C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC  
10,  
PC C12P21/02,C12Q1/68//C12P21/08,C12N15/00,C12N5/00 CC  
Secretory protein or membrane protein

PH	Key	Location/Qualifiers	Location/Qualifiers
FT	CDS		
FEATURES			
source		1. .2563	/organism="Homo sapiens"
			/mol_type="genomic DNA"
			/db_xref="taxon:9606"
ORIGIN			
Alignment Scores:			
Pred. No.:	0	Length:	2563
Score:	433.00	Matches:	733
Percent Similarity:	99.19%	Conservative:	0
Best Local Similarity:	99.19%	Mismatches:	3
Query Match:	58.83%	Indels:	6
DB:	6	Gaps:	0
US-10-045-815-4 (1-736) x BD123548 (1-2563)			
QY	1	MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValValAlaAlaLas	20
DB	31	ATGGCGGTACGCGGTGTAAGCTGTGACACACTGTGGCTCTCGTGGCGGCTGCC	90
QY	21	GlAlaGluValGluSerGluAlaGlyTTPGlyMetValThrProAspLeuLeuPheAla	40
DB	91	CAAGCCGAGGTCGAGTCCGAGGAGGATGGGCGATGGTGGCTGATCTGCTCTCGCC	150
QY	41	GluGlyThrAlaAlaValAlaArgGlyAspTTPProGlyValValLeuSerMetGluArg	60
DB	151	GAGGGACCCAGCCTACGCGCGGGGACTGGCCCGGGGTGGTCTTGAGCATGGACGG	210
QY	61	AlaLeuArgSerArgAlaAlaLeuAlaLeuAlaLeuArgGlyCysArgThrGlnCysAla	80
DB	211	GCCTGGCCCTCCCGGGCAGCCCTCGCGCCCTTCGCGCTGGCTGGCCACCCAGTGTGCC	270
QY	81	AlaAspPheProTTPGluLeuAspProAspTTPSerProSerProAlaGlnAlaSerGly	100
DB	271	GCCGACTTCCCGTGGAGCTGGACCCCGACTGTGTCCCCAGCCCGCCAGGCTCGGGC	330
QY	101	AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLe	120
DB	331	GCCGC-CGCCCTCGCGACCTGAGCTTCTTCGGGGGCTTCTGGTCCGCGCTGCCCT	389
QY	120	uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh	140
DB	390	GCGCGCTGCTCGCGCGCGCGCGCCACTCGCTCAGCGAAGAGATGGAGCTGGAGTT	449
QY	140	eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysLeuAsnLysLeuG	160
DB	450	CCGCAAGCGAGCCCTTACACTACTCTGAGGTCCGCTACTTCAAGATCAACAGTTGGA	509
QY	160	uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetG	180
DB	510	GMAAGCTGTGTCAGCACACACCTTCTTCGTGGGCAATCCTGAGCACATGGAAATGCA	569
QY	180	nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe	200
DB	570	GCAGAACTAGACTATTACCAACCATGTCTGAGTGAAGAGCGCCACTTCAAGGATCT	629
QY	200	uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluG	220
DB	630	TGAGACTCAACCCCATATGCAAGAAATTCGACTGGGAGTCCGACTCTACTCAGAGGAACA	689
QY	220	nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrG	240
DB	690	GCCACAGAAAGCTGTGCCCCACCTTAGAGCGCGCTGCAAGAATACTTGTGGCCCTATGA	749
QY	240	uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr	260
DB	750	GGAGTGGCGTCCCTCTCGGAGGGCCCTATGACTACATGAGTGGCTACACTACCTTGGTA	809
QY	260	rAsnAlaAspLeuPheGlnAlaLeuThrAspHisTyrIleGlnValLeuAsnCysLysG	280

DB	810	CAACGCTGACCTCTTCCAGGCGCATCACAGATCAATTATCATCCAGGTCTCTCAACTGTAAGCA	869
QY	280	nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe	300
DB	870	GAACTGTGTACGAGCTTGTCTCCACCAAGTCGAGAGAGCCCTTTGAAGACTTCTCT	929
QY	300	uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl	320
DB	930	CCCATCGCATTAATAATTATCTGCAGTTTGCCTACTATAACATTGGGAATTATACACAGC	989
QY	320	aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA	340
DB	990	TGT-TGAATGTGCCAAGACCTATCTCTCTTCTCCCAATGACGAGGTGATGAACCAAA	1048
QY	340	snLeuAlaTyr-TyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG	360
DB	1049	ATTGGCCCTATTATGCAGCTATCTTGGAGAAGAACACACAGATCCATCGGCCCCCGTG	1108
QY	360	luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaT	380
DB	1109	AGAGTGCACAGAGTACCGACAGCGAAGCTACTGGAAAAAGAACTGCTTTCTTCGCTT	1168
QY	380	yrAspValPheGlyTleProPheValAspProAspSerTTPThrProGluGluValIleP	400
DB	1169	ATGATGTTTTTGGAAATCCCTTTTGGATCCGGATTCTATGACCTCCAGAGAGTGTTC	1228
QY	400	toLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG	420
DB	1229	CCAAGAGATTGCAAGAGAAACAGAAAGTCAGAACCGGAAACAGCCGTACGCATCTCCAGG	1288
QY	420	luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL	440
DB	1289	AGATTGGGAACCTTATGAAGAAATCGAGACCCCTGTGGAAGAGAAGAACCAAGGAGTCA	1348
QY	440	euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuT	460
DB	1349	TGGATGTGACGAGACTGACCCGGGAGGTGGCCCTCTGCTGTATGAGGCATCACTCA	1408
QY	460	hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer	479
DB	1409	CCATGAACCTCCAAACTCCTCTGAATGGTTC-CCAGCGGGTGGTGGATGGAGCGGTAATCTCT	1467
QY	480	AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly	499
DB	1469	GACCAAGAGTGTGAGAGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGC	1527
QY	500	TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe	519
DB	1528	TACCGGGTTCAGACTCCCACTCTCCCAATGAAAAAGTTCTATGTGTCTCTCTCTTC	1587
QY	520	LysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyr	539
DB	1588	AAAGCCCTCAAGCTGGGCAAGAGCAAGTTCCTCTGCAGAGTGCCACCTGTACTTAC	1647
QY	540	AsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeu	559
DB	1648	AACGTGACGGAAGAGTGGCGGCATCATGGAGTCTCTACTTCCGCTGGATACGCCCTTC	1707
QY	560	TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnIleGluArg	579
DB	1708	TACTTTTCT	1767
QY	580	LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal	599
DB	1768	AAGAGTATAGTATCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGT	1827
QY	600	CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn	619
DB	1828	TGTGTCAAAGAGCCCGGAGCTTACACCTTCCGCGACTACAGGCCATCTCTTTTACCTAAAT	1887
QY	620	GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla	639
DB	1888	GGGAGCTTCGATGGCGGAAACTTTTATTCTACTGAATGGATGCCAGACCGTACGCA	1947

```

QY      640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659
DB      1948 GAGGTGCAGCTCAGTGTGGAGAGCGCGTGGGATCTCTTTCAGGACACTGAAACCCACAT 2007
QY      660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTTPPheThrLeuAsp 679
DB      2008 GGAGTGAAGGTGTTCACACAGGGGAGCGCTGTGCCATCGCCCTGTGTTTCAACCTGGAC 2067
QY      680 ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer 699
DB      2068 CCTGCACACAGCGAGCGGACAGGGTGCAGGACAGTACCTGTGTGAAGATGCTCTTCAGC 2127
QY      700 ProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGlu 719
DB      2128 CCAGAGAGATGGACCTCTCCAGAGAGAGCCCTGTGATGCCACAGGCGGCCCCCGGAA 2187
QY      720 ProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736
DB      2188 CCTGCACAAGATCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2238

RESULT 4
AK075418      2563 bp      mRNA      linear      PRI 03-SEP-2002
LOCUS      Homo sapiens cDNA PSECO109 fis, clone NT2RP2005671, moderately
DEFINITION      similar to Rattus norvegicus leprecan (leprel) mRNA.
ACCESSION      AK075418
VERSION      AK075418.1 GI:22761493
KEYWORDS      oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      1
AUTHORS      Ota, T.; Nishikawa, T.; Suzuki, Y.; Kawai-Hio, Y.; Hayaishi, K.;
      Ishii, S.; Saito, K.; Yamamoto, J.; Wakamatsu, A.; Nagai, T.;
      Nakamura, Y.; Nagahara, K.; Sugano, S. and Isogai, T.
      HRI human cDNA sequencing project
      Unpublished
      2 (bases 1 to 2563)
      Isogai, T. and Yamamoto, J.
      Direct Submission
      Submitted (20-MAR-2002) Takao Isogai, Helix Research Institute,
      Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
      (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
      HRI human cDNA sequencing project; cDNA 5' - & 3' - end one pass
      sequencing, clone selection and full insert sequencing; Helix
      Research Institute (supported by Japan Key Technology Center etc.);
      cDNA library construction; Institute of Medical Science, University
      of Tokyo, Laboratory of Genome Structure, Human Genome Center.
      Location/Qualifiers
      1. .2563
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="NT2RP2005671"
      /cell_line="NI2"
      /cell_type="teratocarcinoma"
      /clone_lib="NT2RP2"
      /note="cloning vector: pME18SPL3-mRNA from NT2 neuronal
      precursor cells after 2-weeks retinoic acid (RA)
      induction"
      31. .2241
      /note="unnamed protein product"
      /codon_start=1
      /protein_id="BAC11608.1"
      /db_xref="GI:22761494"
      /translations="MAVRALKLTLVAVVAASQAEVSEAGWGMVTDLLFAEGTA
      AYARGDPVGVLSMRALRLRRCQCAADFPWELDPDPSPAAGAA
      ALRDLSPFGGLRRACLRLCPGAHSLSEMELEFSPYNYLOVAYKINKLE
      KAVAAHTFVCGHPENEMQNLDYQYQTSVKVEADFKOLETPHNPQFRLGVLYSE
      EQPOEAVPHLEALQYFYAYEACRLCEGYDYGNYLEYNADFOAITDHYIOVL
      NCKQNCVTELASHSPREKPEDFLPHYNYLOPAYNYIGNYTOAVECAKTYLLFPEND

```

```

EVMNQNLAYAAMLGEBHTRISIGPRESAKYRQSRLLLEKELLFFAYDVFGIPFVDPDPS
WTPBEVTPRIQEKQKRSERATVRSIQEINLMKEIETLVEBKTESLDVRLRREGG
PLYEGISLTWNKSLNGSRQVMDGVISDECEQLRLTNVAATSGDGYGQTSPT
PNEKFGYTVTFKALKGQGVKVPDQSAHLTYNTVEKVRRIIMESYDITDPLTFYSGSHL
VCRTEIEVQAEKDDSHPDNCILNAETLVCVKPEPFTFRDYSAILYLNGDFG
GNFYTELDATKIVAEVOQCGRAVGFSSGTENPHGVKAVTRGORCAIALAFTLDPRH
SERDRVQADDLVKMLFSPEMDLSQEQPLDAQQGPPEPAQBSLSGSSESKPKDEL"

ORIGIN
Alignment Scores:      0      Length:      2563
Pred. No.:      433.00      Matches:      733
Score:      99.19%      Conservative:      0
Percent Similarity:      99.19%      Mismatches:      3
Best Local Similarity:      99.19%      Indels:      6
Query Match:      58.83%      Gaps:      0
DB:      9

US-10-045-815-4 (1-736) x AK075418 (1-2563)

QY      1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaAlaSer 20
DB      31 ATGGCGGTACGCGCGTGAAGCTGCTGACACACACTGCTGGCTGCTGCGCGCTGCCCTCC 90
QY      21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40
DB      91 CAAAGCCGAGGTCCAGTCCGAGGAGGATGGGCGATGGTGAAGCCCTGATCTGCTCTTCCGC 150
QY      41 GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60
DB      151 GAGGGGACCGGACCTACGCGCGGGGACTGGCCCGGGTGGTCTCTGAGCATGGAACGG 210
QY      61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
DB      211 GCGCTGGCTCCCGGGGAGCCCTCGCGCCCTTGGCTGGCTGGCCACCCAGTGTGCC 270
QY      81 AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100
DB      271 GCGGACTTCCGCTGGGAGCTGACCCGACTGGTCCCCCAGCCGCGCCAGCCCTCGGCC 330
QY      101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLe 120
DB      331 GCGGC -CGCCCTGCGCGACCTGAGCTTCTCGGGGCGCTTCTGGCTGCGCTGCTGCT 389
QY      120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140
DB      390 GCGCGCTGCTCGCGCGCGCGCGCCACTCGCTCAGCGAAGAGATGGAGCTGGAGTT 449
QY      140 eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuG1 160
DB      450 CCGCAAGCGGAGCCCTCAACTACCTGCGAGTGGCTTCAAGATCAACAGTGGGA 509
QY      160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyLysnProGluHisMetGluMetG1 180
DB      510 GAAAGCTGTGCTGCAGCACACACTTCTTCTGGGCAATCTCTGAGCACATGGAATGCA 569
QY      180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200
DB      570 GCAGAACCTAGACTATATTACCAAAACCATGCTCGGAGTGAAGGAGGCGCCACTTCAAGGATCT 629
QY      200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluG1 220
DB      630 TGAGACTCAACCCCATATGCAAGAAATTTGCACTGGGAGTGGGACTCTACTCAGAGGAACA 689
QY      220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrG1 240
DB      690 GCCACAGGAAGCTGTGCCCCACCTAGAGGCGCGCTGCAAGAATACTTTGTGGGCTATGA 749
QY      240 uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260
DB      750 GGAGTGGCTGCCCTCTCGGAGGCGCCCTATGACTACGATGGCTACAACTACCTTGGATGA 809
QY      260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysG1 280

```

810 CAACGGTACCTCTTCCAGGCCATCAGATCATTTACATCCAGGTCTCTCAACTGTAAACA 869  
280 nAnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
870 GAACTGTGTCCAGGAGCTTCTCCACCAAGTCGAGAGAGCCCTTTGAAGACTTCT 929  
300 uProSerHisTyAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320  
930 CCATCGCATTAATATCTGAGTTTGCCTACTATATACATTGGGAATATATACACAGGC 989  
320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
990 TGT-TGAATGTGCAAGACCTATCTTCTCTTCCCAATGACGAGGTGATGAACCAAA 1048  
340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluHisThrArgSerIleGlyProArgG 360  
1049 ATTGGCCTAATATGACCTATCTTGGAGAGAACACACAGATCAATCGGCCCCCGTG 1108  
360 luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuPhePheAla 380  
1109 AGAGTGCACAGGAGTACCGACAGCGAAGCCTACTGGAAAAGAACTGCTTTTCTCGCTT 1168  
380 YrAspValPheGlyIleProPheValAspProAspSerThrProGluGluValIleP 400  
1169 ATGATGTTTTTGGAAATTCCTTTTGGATCCGATTCTGGAATCCAGAGAGAGTGAATC 1228  
400 roLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
1229 CCAAGAGATTGCAAGAGAACAGAGTCCAGACGGGAAACAGCCGTACGCATCTCCAGG 1288  
420 luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440  
1289 AGATTGGGAACCTTATGAAGAAATCGAGACCCCTTGTGGAAGAGAAAGCAAGAGTCA 1348  
440 euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuT 460  
1349 TGAATGTGACACACTGACCCCGGAGGTGGCCCTCTGTTATGAGGCATCGTCTCA 1408  
460 hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer 479  
1409 CCATGAACCTCCAACTCTCTGAATGGTTC-CCAGCGGGTGTGTGATGACGGCGTAATCTCT 1467  
480 AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499  
1468 GACCACAGGTGTGAGGAGTGCAGAGACTGACCAATGTGCGACCAACCTCAGAGAGTGC 1527  
500 TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe 519  
1528 TACGGGGTTCAGACCTCCCCACATACTCCCAATGAAAGTTCTATGGTGTCACTGTCTTC 1587  
520 LysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyr 539  
1588 AAAAGCCCTCAAGCTGGGGCAAGAGGCAAAAGTTCCCTGTGAGAGTGCCACACCTGTACTAC 1647  
540 AsnValThrGluLysValArgGlyLeMetGluSerTyrPheArgLeuAspThrProLeu 559  
1648 AAGTGTGAGAGAGGTGGCGGCATCATGAGTCTTCTTCCGCTTGGATACGCCCTTC 1707  
560 TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArg 579  
1708 TACTTTTCTTACTCTCTATCTGTTGCGGCACCTGCCATCGAAGAGTCCAGGCAGAGAG 1767  
580 LysAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599  
1768 AAGATGATAGTATCATCGTCCAGTCCAGTGAACATCATCTGAATCCGAGACCCCTCGTG 1827  
600 CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619  
1828 TGTGTCAAGAGCCCCCAGCCTACACCTTCCGCGACTACAGCGCCATCTTTTACCTAAAT 1887  
620 GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla 639  
1888 GGGGACTTCGATGCGGAAACTTTTATTTTCACTGAATGGATGCCAAGACCGTGACGGCA 1947

QY 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659  
Db 1948 GAGGTGACGCTTCAGTGTGGAAGAGCCGTGGATTCTCTTCAGGCACCTGAAACCCCAT 2007  
QY 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAsp 679  
Db 2008 GGAGTGAAGGTGTCCACAGGGGGCAGCGCTGTGCATCGCCCTGTGGTTCCACCTGCAG 2067  
QY 680 ProArgHisSerGluArgAspArgValGlnAlaAspAlaAspLeuValLysMetLeuPheSer 699  
Db 2068 CCTGACACAGCGAGCGGAGCAGGGTGCAGCAGATGACCTGGTGAAGATGCTCTTCAGC 2127  
QY 700 ProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGlnGlnGlyProProGlu 719  
Db 2128 CCAGAGAGATGGACCTCTCCAGGAGCAGCCCTGGATGCCAGCGGCCCGCCCGAA 2187  
QY 720 ProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2188 CCTGCACAGAGTCTCTCTCAGCAGTGAATCGAAGCCCAAGGATGAGCTA 2238

## RESULT 5

AX704765 2593 bp DNA linear PAT 04-APR-2003  
LOCUS Sequence 69 from Patent WO0202634.  
DEFINITION AX704765  
ACCESSION AX704765  
VERSION AX704765.1 GI:29561431

KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens

## REFERENCE

AUTHORS Gururajan,R., Hafalia,A.J., Kallick,D.A., Patterson,C., Azimzai,Y.,  
Khan,F.A., Xu,Y., Yao,M.G., Yue,H., Au-Young,J., Batza,S.,  
Baughn,M.B., Borowsky,E.A., Lo,T.P., Lu,D.A., Lu,Y., Tang,Y.T.,  
Yang,J., Zingler,K.A., Deleane,A.M., Gietzen,K., Marcus,G.A.,  
Nguyen,D.B., Policky,J.D., Ramkumar,J., Thangavelu,K., Walla,N.K.,  
and Warren,B.A.  
TITLE Human extracellular matrix and cell adhesion polypeptides  
JOURNAL Patent: WO 0202634-A 69 10-JAN-2002;  
INCYTE GENOMICS INC. (US)

## FEATURES

## source

1..2593 Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 6780147CB1"

## ORIGIN

Alignment Scores:  
Pred. No.: 0 Length: 2593  
Score: 433.00 Matches: 733  
Percent Similarity: 99.19% Conservative: 0  
Best Local Similarity: 99.19% Mismatches: 3  
Query Match: 58.83% Indels: 6  
DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x AX704765 (1-2593)

QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
Db 42 ATGGCGGTACGGCGCTTGAAGCTGTGACACACACTGTGGCTGTGGTGGCGCTCC 101  
QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuPheAla 40  
Db 102 CAAGCCGAGGTGAGTCCGAGCAGGATGGGCGATGGTGAACCTGATCTGCTCTGCC 161  
QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
Db 162 GAGCGGACCGGAGCTACGCGCGGGGACTGGCCCGGGGTGGTCTCTGAGATGGAACGG 221  
QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80



Db 222 CGCTGGCTCCGGGAGCCCTCGGGCCCTCGCTGGCTGGCCGACCCAGTGTGCC 281  
Qy 81 AlaSpPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly 100  
Db 282 GCCGACTTCCCGTGGAGCTGGACCCGAGCTGGTCCCGGACCCCGGCTCCGGC 341  
Qy 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyLeuLeuArgAlaAlaCyste 120  
Db 342 GCCGC-CGCCCTGGCGACCTGAGCTTCTTCGGGGCCCTTCGGCTCGCCCTGCCCT 400  
Qy 120 uArgCysLeuGlyProProAlaAlaHisSerLeuSerGluMetGluLeuGluPh 140  
Db 401 GCGCGCTGCCGCGCGCGCGCCCACTCGCTCAGCGAGAGATGGAGTT 460  
Qy 140 eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuG 160  
Db 461 CCGAAGCGAGCCCTACAACTACCTGCAGTCCCTTCAAGATCAACAAGTTGGA 520  
Qy 160 uLysAlaValAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetG 180  
Db 521 GAAAGCTGTCTCGAGCACACACTTCTTCGTGGGCAATCCTGAGCACATGGAAATGCA 580  
Qy 180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
Db 581 CGAAGAACCTAGACTATTACCAACCATGTCTGGAGTGAAGAGCGCCGACTTCAAGGATCT 640  
Qy 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluG 220  
Db 641 TGAGACTCAACCCCATATGCAAGAAATTCGACTGGAGTGGAGTCTACTTCAGAGGAC 700  
Qy 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrG 240  
Db 701 GCCACAGAGAGCTGTGCCCACTAGAGCGCGCTGCAAGATATCTTTGTGGCCTATGA 760  
Qy 240 uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTy 260  
Db 761 GGAGTGGCGTCCCTCTCGAAGGGCCCTATGACTAGTGGCTACCACTACCTTTGAGTA 820  
Qy 260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysG 280  
Db 821 CAACGCTGACCTTCCAGGCCCATCACAGATCATATCATCCAGTCTCCAACTGTAAACA 880  
Qy 280 nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
Db 881 GAACCTGTCTCGAGCTTGTCTCCCAACCAAGTCGAGAGAGCCCTTTGAAGACTTCCT 940  
Qy 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320  
Db 941 CCCATCGCATTAATATATCTGCAAGTTGCTACTATATCAATTGGGAATATACACAGGC 1000  
Qy 320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
Db 1001 TGT-TGAATGTGCCAAGACCTATCTTCTTCTTCCCAATGACGAGGTGATGAACCAAA 1059  
Qy 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArg 360  
Db 1060 ATTTGGCTATTATGCACTATGTTTGAGAGAACACACAGATCCATCGGCCCCCGTG 1119  
Qy 360 luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlar 380  
Db 1120 AGAGTGCAAGAGTACCGACAGCGAGCCCTACTGGAAAAGAACTGCTTTTCTCGCTT 1179  
Qy 380 yrAspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIleP 400  
Db 1180 ATGATGTTTTGGAAATTCCTTTTGGATCCGGATTATGACTTCCAGAAAGATGATTC 1239  
Qy 400 roLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
Db 1240 CCAAGAGATTGCAAGAGAAACAGAAAGTCAGAACGGGAAACAGCCGTACGCATCTCCAGG 1299  
Qy 420 luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440

Db 1300 AGATTGGGAACCTTATGAAGGAATTCGAGACCCCTTGTGGAAGAGAACCAAGAGTCCAC 1359  
Qy 440 euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuT 460  
Db 1360 TCGATGTGAGCACTGACCCCGGAAGTGGCCCTCTGTATGAAGGCATCAGTCTCA 1419  
Qy 460 hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer 479  
Db 1420 CCATGAACCTCCAACTCTCTGAATGTTTCC-CCAGCGGTGGTGGATGGAGCGGTAACTCT 1478  
Qy 480 AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499  
Db 1479 GACCAAGAGTGTGAGAGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGC 1538  
Qy 500 TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe 519  
Db 1539 TACCGGGTGCAGACCTCCCACTACTCCCAATGAAAAGTTCTATGGTGTCTACTGTCTTC 1598  
Qy 520 LysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyr 539  
Db 1599 AAAGCCCTCAAGCTGGGCAAGAGCAAGTTCCTCTGCAGAGTGCCTCTGACTAC 1658  
Qy 540 AsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeu 559  
Db 1659 AACGTGACGAGAGAGGTGGCGCATCATGAGTCTTCTTCCGCTGGATACGCCCTC 1718  
Qy 560 TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArg 579  
Db 1719 TACTTTCTCTCTCATCTGTGTGGCGCACTGCATCGAAGAGTCCAGCAGAGAGG 1778  
Qy 580 LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599  
Db 1779 AAGGATGATAGTATCCAGTCCAGTCCAGCAACCTGCATCCTGAATGCCGAGACCTGTG 1838  
Qy 600 CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619  
Db 1839 TGTGTCAAGAGCCCCAGCTTACACTTCGCGACTACAGGCACTCTTTACTTAAT 1898  
Qy 620 GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla 639  
Db 1899 GGGGACTTCGATGGCGAAACTTTTATTTTCACTGAATCTGGATGCCAAGCCGTGACGCA 1958  
Qy 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659  
Db 1959 GAGGTGACCTCTGATGGAGAGCCGTGGATTTCTTTCAGGCATGAAAACCCACAT 2018  
Qy 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAsp 679  
Db 2019 GGAGTGAAGCTGTCAACAGGGGCGAGCTGTGCTCATCGCCCTGTGGTTTCACTGGAC 2078  
Qy 680 ProArgHisSerGluArgAspArgValGlnAlaAspLeuValLysMetLeuPheSer 699  
Db 2079 CTTCCACACAGCAGCGGAGGTGAGGAGTGCAGGAGTACCTGGTGAAGATGCTCTTCAGC 2138  
Qy 700 ProGluGluMetAspLeuSerGlnGluProLeuAspAlaGlnGlnGlyProProGlu 719  
Db 2139 CAGAAGAGATGAGCTTCTCCAGAGGAGCCCTGGATGCCAGAGGGCCCCCGAA 2198  
Qy 720 ProAlaGlnGluSerLeuSerGlySerGlySerLysProLysAspGluLeu 736  
Db 2199 CCTGCACAGAGTCTCTCTCAGGCAAGTGAATCGAAGCCCAAGAGTGAAGCTA 2249

## RESULT 6

AX877288  
LOCUS AX877288 2595 bp DNA linear PAT 17-DEC-2003  
DEFINITION Sequence 12193 from Patent EP074617.  
ACCESSION AX877288  
VERSION AX877288.1 GI:40032024  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.



REFERENCE 1  
 AUTHORS Ota, T., Isozaki, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,  
 Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.  
 TITLE Primers for synthesizing full-length cDNA and their use  
 JOURNAL Patent: EP 1074617-A 12193 07-FEB-2001;  
 Research Association for Biotechnology (JP)  
 FEATURES  
 source 1. .2585  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 53. .2263  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CA289658.1"  
 /db\_xref="GI:40032025"  
 /translation="MAVRALKLLTLLAVAAASQAEVSEACGMVTPDLLFAECTA  
 AYARGWPQVILSMERALSRAALRALRLRCRTOCAADFPWELDPWSFSPAQSGAA  
 ALKDLFFGLLRRAACLRCLGPPAAHSLSEMELEFRKRSPYNTLQVAYFKINKLE  
 KAVAAHFFVGVNEMEQNDLYYTQWTSVKEADFKOLETPQHPQEFRLGVRLYSE  
 EQPQEAHPLEALAQEFYVAYEACRALCEGPDYDGYNYLEYNADLFQAITDHYIOVL  
 NCKQNCVTELASHPSEKFEDELPHSHYNLQFAYVYNYGNVTOAVECAKTYLLFPND  
 EVNNQNLVYAAMLBEHTRSIGPRESAKYRORSLEKELLFFAYDVFQIPFVPPDS  
 WTEPEVLPKELQKQKSERETAVLSQELGNLMKELETLVESKTESLDVSLTRGG  
 PLYEGISLTMSKLNGSORVNDGVISDHCEQELQRLTNVAISGDSYRQTSPT  
 PNEKFGVTVFKALKIGQEGKVPLOSAHLIYNTKVRIMESYFRDLTPLYFSYHL  
 VCFRTALVEOVAERKDSHPVHVDNCLINAEITLVCKEPPAYTFRDYSAILYLNGFDG  
 GNFYFTELDKTYTAEVQPOCGRAVGFSSGTENPHGVKAVTRQRCALIAWFTLDPRH  
 SERDRVQADDLVXKMLFSPEDMLSQEPLDAQQPPPEAQSLESSESXPKDEL"

ORIGIN

Alignment Scores:  
 Pred. No.: 0 Length: 2585  
 Score: 433.00 Matches: 733  
 Percent Similarity: 99.19% Conservative: 0  
 Best Local Similarity: 99.19% Mismatches: 3  
 Query Match: 58.83% Indels: 6  
 DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x AX877288 (1-2585)

QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValValAlaAlaLaser 20  
 DB 53 ATGGCGGTACGGCGTTGAGAGTGTCTGACACACAGTGTGCTGCTGCGCGCTGCTCC 112  
 QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40  
 DB 113 CAAGCCGAGGTCTGAGTCCGAGGAGGATGGGCGATGGTGTGACGCTGATCTGCTCTTCGCC 172  
 QY 41 GluGlyThrAlaAlaValAlaValArgAlaValValValValValValValValValVal 60  
 DB 173 GAGGGGACCGAGCTACGCGCGGGGAGTGGCGCGGGGTGCTCTGAGCATGGAGCGG 232  
 QY 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
 DB 233 CGCGTGGCTCCCGGGCAGCCCTCCGCGCTTCCGCTGCGCTGCGCGACCCAGTGTGCC 292  
 QY 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProAlaGlnAlaSerGly 100  
 DB 293 GCCCACTTCCCGTGGGAGCTGAGACCCGACTGTGTCCTCCAGCCCGCGCGCTCGGGC 352  
 QY 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLe 120  
 DB 353 GCCGC-CGCCCTGGCGACCTGAGCTTCTTCGGGGGCTTCTGCTGCGCTGCTGCTGCT 411  
 QY 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140  
 DB 412 CGCGCGCTGCTCGCGCGCGCGCGCCACTGCTGCTCAGCGAAGAGATGGAGCTGGAGTT 471  
 QY 140 eArgLysArgSerProTrpAsnTrpLeuGlnValAlaValPheLysLysLysLeuGlu 160  
 DB 472 CCGCAAGCGGAGCCCTTACAACTACCTGAGCTGCGCTTCTTCAAGATCAACAGTTGGA 531

QY 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGlu 180  
 DB 532 GAAAGCTGTGCTGCGAGCACACACACCTTCTCGTGGCAATCTCTGAGCACATGGAATGCA 591  
 QY 180 nGlnAsnLeuAspTrpTrpGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
 DB 592 SCAGAACCTAGACTATTACCAAAACATGTCTGGAGTGAAGGAGCGCGACTTCAAGAGTCT 651  
 QY 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTrpSerGluGluGlu 220  
 DB 652 TGAGACTCAACCCCATATGCAAGAAATTTGAGTGGAGTGGGACTCTACTCAGAGGAACA 711  
 QY 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTrpPheValAlaTrpGlu 240  
 DB 712 GCCACAGGAAGCTGTGCCCCACCTAGAGCGCGCTGCAAGATACTTTGTGGCCTATGA 771  
 QY 240 uGluCysArgAlaLeuCysGluGlyProTrpAspTrpAspGlyTrpAsnTrpLeuGluTrp 260  
 DB 772 GGAGTGGCTGCCCTCTCGGAGGCGCTTATGACTACGTGGTACACTACTCTTGAGTA 831  
 QY 260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTrpIleGlnValLeuAsnCysLysGlu 280  
 DB 832 CAACGCTGACCTCTTCCAGGCGCATCACAGATCATTACATCCAGGTCTCTCAACTGTGAACA 891  
 QY 280 rAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
 DB 892 GAACGTGTGTCGAGGCTGTCTTCCCAAGTCGAGAGAGCCCTTTGAAGACTTCT 951  
 QY 300 uProSerHisTrpAsnTrpLeuGlnPheAlaTrpTrpAsnIleGlyAsnTrpThrGlnAla 320  
 DB 952 CCCATCGCATTAATATCTGCGCTTCTGCTACTATAACATTTGGAAATATACACAGGC 1011  
 QY 320 aGly-GluCysAlaLysThrTrpLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
 DB 1012 TGT-TGAATGTGCCAAGACCTATCTTCTTCTTCCCAATGACGAGGTGATGAACCAAA 1070  
 QY 340 snLeuAlaTrpTrpAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG 360  
 DB 1071 ATTGGCTTATTATGACGATGCTTGGAGAGAACACACACAGATCCATCGGCCCCCGTG 1130  
 QY 360 luSerAlaLysGluTrpArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaTr 380  
 DB 1131 AGAGTGCCAAAGAGTACCGACAGCGAAGCTACTTGAAAAAGAACTGCTTTTCTTCGGCT 1190  
 QY 380 yrAspValPheGlyIleProPheValAspProAspSerTrpThrProGluGluValIleP 400  
 DB 1191 ATGATGTTTTTGGAAATCCCTTTTGGATCCGATTCGATTCATGACCTCCAGAGAGTATTC 1250  
 QY 400 rLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
 DB 1251 CCAAGAGATTGCAAGAGAAAACAGAACTCAGAAACGGGAAACAGCCGTACCCATCTCCAGG 1310  
 QY 420 luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440  
 DB 1311 AGATTTGGGAACCTTATGAAGAAATCGAGACCTTTGTGGAAGAGAACCAAGGAGTGCAC 1370  
 QY 440 euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTrpGlyIleSerLeuTr 460  
 DB 1371 TGGATGTGAGCAGACTGACCCGGGAGGTGGGCCCTCTGCTGATGAGGCATCAGTCTCA 1430  
 QY 460 hrMetAsnSerLysLeuLeuAsnGlyTrp-GlnArgValValMetAspGlyValIleSer 479  
 DB 1431 CCATGAACCTCCAAACTCTCTGAATGGTTC-CCAGCGGGTGGTGGATGGACGGCGTAATCTCT 1489  
 QY 480 AspHisGluCysGlnGluLeuGlnAlaArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499  
 DB 1490 GACACGAGTGTGAGAGCTGCAGAGCTGACAGCTGACCAATGTGGCAGCAACCTCAGGAGTGGC 1549  
 QY 500 TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTrpGlyValThrValPhe 519  
 DB 1550 TACCGGGGTGACAGCTCCCCACATCTCCCAATGAAAAGTCTCTATGGTGTCACTGTCTTC 1609  
 QY 520 LysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTrpTyr 539



832 CAACGGTCAGCTCTTCCAGCCCATCACAGATCATTTACATCCAGGTCTCTCAACTGTAAACA 891  
Qy 280 nAnCysValThrCluLeuAlaSerHisProSerArgGluLysProPheLeuAspPheLe 300  
Db 892 GAACGTGTTCAGGAGCTTCTCCACCACCAAGTCGAGAGAGCCCTTTGAAGACTTCTT 951  
Qy 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320  
Db 952 CCCATCGCATTAATAATTCTGCAGTTTGTCTACTATAACATTGGGAATTATACACAGGC 1011  
Qy 320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
Db 1012 TGT-TGAATGTGCAAGACCTATCTCTCTCTTCCCATGATGAGAGGTGATGAACAA 1070  
Qy 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluLysHisThrArgSerIleGlyProArg 360  
Db 1071 ATTTGGCTTATTATGAGCTATGCTTGGAGAAGACACACACAGATCCATCGGCCCGCTG 1130  
Qy 360 luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAla 380  
Db 1131 AGAGTGCACAGGAGTACCGACAGCGAGCCCTACTGGAAAAAGAACTGCTTTTCTTCGCTT 1190  
Qy 380 YrAspValPheGlyIleProPheValAspProAspSerTrrThrProGluGluValIleP 400  
Db 1191 ATGATGTTTTTGGAAATTCCTTTTGTGGATCCGATTCATGATCTCAGAAAGATGATTC 1250  
Qy 400 roLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
Db 1251 CCAAGAGATTGCAAGAGAAACAGAAAGTCAGAACGGGAAACAGCGGTACGCATCTCCAGG 1310  
Qy 420 lulleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440  
Db 1311 AGATTGGGAACCTTATGAAGAAATCGAGACCCCTTGTGGAGAGAGAACCAAGAGGTAC 1370  
Qy 440 euAspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyIleSerLeu 460  
Db 1371 TGAATGTGACAGACTGACCGCGGAGGTGCGCCCTGCTGTATGAAGGATCAGTCTCA 1430  
Qy 460 hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer 479  
Db 1431 CCATGAACCTCAAACTCCTGAATGGTTC-CCAGCGGTGTGTGATGACGCGCTAATCTCT 1489  
Qy 480 AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499  
Db 1490 GACCACAGGTGTACGAGGTGCGAGACTGACCAATGTGCAGCAACCTCAGAGATGGC 1549  
Qy 500 TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe 519  
Db 1550 TACCGGGTTCAGACCTCCCCACATACCTCCCAATGAAGAAGTTCTATGGTGTCACTGCTTC 1609  
Qy 520 LysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyrTyr 539  
Db 1610 AAAGCCCTCAAGCTGGGGCAAGAGGCAAGTTCTCTGAGAGTGGCCACCTGTACTAC 1669  
Qy 540 AsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeu 559  
Db 1670 AACGTGACGAGAGGTGCGGCGCATCATGAGTCTTACTTCCGCTGGATACGCCCTC 1729  
Qy 560 TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArg 579  
Db 1730 TACTTTTCTACTCTCATCTGTGTGCGCACTGCCATCGAAGAGGTCACGAGGAGAGG 1789  
Qy 580 LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599  
Db 1790 AAGGATGATAGTATCCAGTCCACGTGACACTGTGATCTGATGCGGAGACCTCGTG 1849  
Qy 600 CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619  
Db 1850 TGTGTCAAAGAGGCCCCAGCGCTACACCTTCGCGACTACAGCGGCACTCTTTACCTAAAT 1909  
Qy 620 GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValAla 639  
Db 1910 GGGGACTCGATGCGGAAACTTTTATTCTACTGAATGGATGCAAGACCGTGACGGCA 1969

Qy 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659  
Db 1970 GAGGTGAGCCCTCAGTGTGGAGAGCCGTGGGATTCCTTTCAGGCACTGAAACCCACAT 2029  
Qy 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAsp 679  
Db 2030 GGAGTGAAGGCTGTCCACAGGGGGCAGCGCTGTGCATCGCCCTGTGGTTTACCCTGGAC 2089  
Qy 680 ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer 699  
Db 2090 CTTCCACACAGCGGCGGACAGGGTGCAGGAGATGACCTGTGTGAAGATGCTCTTCAGC 2149  
Qy 700 ProGluGluMetAspLeuSerGlnGlnProLeuAspAlaGlnGlnGlyProProGlu 719  
Db 2150 CCAGAAGAGATGAGACTCTCTCAGGAGCAGCGCTGTGATGCCAGCAGGGCCCCCGCAA 2209  
Qy 720 ProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAspGluLeu 736  
Db 2210 CTTGCACAAGAGTCTCTCTCAGGAGTGAATCGAAGCCCAAGGATGAGCTA 2260

RESULT 8  
AK027697 2585 bp mRNA linear PRI 01-AUG-2002  
LOCUS Homo sapiens cDNA FLJ14791 fis, clone NT2RP4001064, weakly similar  
DEFINITION to SYNAPTONEMAL COMPLEX PROTEIN SC65.  
AK027697  
ACCESSION AK027697.1 GI:14042570  
VERSION oligo capping; fis (full insert sequence).  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1  
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,  
Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y.,  
Matsumawa,H., Ishii,S.; Kawai,Y., Saito,K., Yamamoto,J.,  
Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Sasaki,N.  
NEDO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 2585)  
Isogai,T. and Otsuki,T.  
Direct Submission  
TITLE Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu Chiba 292-0812, Japan  
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing;  
Research Association for Biotechnology; cDNA library construction,  
5'- & 3'-end one pass sequencing and clone selection; Helix  
Research Institute (supported by Japan Key Technology Center etc.)  
and Department of Virology, Institute of Medical Science,  
University of Tokyo.  
Location/Qualifiers  
1. .2585  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2RP4001064"  
/cell\_line="NT2"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2RP4"  
/note="cloning vector: pME18SFL3-mRNA from NT2 neuronal  
precursor cells after 2-weeks retinoic acid (RA)  
induction."  
53..2263  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAB55305.1"  
/db\_xref="GI:14042571"

CDS  
translation="MAVRALKLTLTLLAVAAASAEVSEBAGWMTVDLLFAEGTA  
AYARGDWGVLSMERALRSRAALPALRLRCTCAADPMPHLPDQSPSPQAQSGAA  
ALRDLUSFFGLLRRAACLRRLCGPPPAISLSEMELEFRKSPNYLQVAFKINKLE

KAVAAATFFVFCNPEHMEMQONLDYQYQMSGVKEADFKDLETOFHMOSFRLGVRLXSE  
 EPOEAVPHLEAAAOEYFVAEECRALCEBGYDYDGNVLEYNADLFQAITDHYIQVL  
 NCKQCVTLASHSRKPFDFLPFSHYNQLOFAYNIGNYQVAECATKYLFFPND  
 EVNQNLAYAAAGEHRETSIGPESAKYRQSRLEKELFFAYDVFQFPFDPDS  
 WTPBEVIPKRLQEKOKSERETAISOIGNLKMEITLVEEATKESLDVSRLTRGG  
 PLLYEGSLTWNLSKLLGSRVNDGVISDHECOELORLTNVAATSDGVYRGOTSPHT  
 PNEKPYGVTFKALKGQEKVPLQSAHLVYNTKRYRIMESYFRDLTDPYFSYSHL  
 VCRFAIEVQAKRDKSDRPHVDNCILNAETLVCVKBPAYTFRDYSAILYLNGDSDG  
 GNIFYTEDAKVTAEVQPCGRAVGFSSGTENPHGVKAVTRGRCRAIALWFLDPRH  
 SERDRVQADDILVKMLFSPPEMDLSQEPDLQAQGPPEPAQESLSGSESKEPDEL"

## ORIGIN

## Alignment Scores:

Pred. No.: 0 Length: 2585  
 Score: 433.00 Matches: 733  
 Percent Similarity: 99.19% Conservative: 0  
 Best Local Similarity: 99.19% Mismatches: 3  
 Query Match: 58.83% Indels: 6  
 DB: 9 Gaps: 0

US-10-045-815-4 (1-736) x AK027697 (1-2585)

QY	260	rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysG	280
Db	832	CAACGCTGACCTCTTCCAGCGCATCACAGATCATTTACATCCAGGTCTCTCAATGTGAAGCA	891
QY	280	nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe	300
Db	892	GAACGTGTACGGAGCTTGTCTCCACCAGTCAGAGAGACCTTTGAAGACTTCTCT	951
QY	300	uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl	320
Db	952	CCATCGCATTAATAATATCTGCAGTTTGCCTACTATAACATTTGGAATTTATACAGGC	1011
QY	320	aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA	340
Db	1012	TGT-TGAATGTGCCAAGACCTATCTCTCTCTCTCCCAATGACGAGGTGATGAACCAA	1070
QY	340	snLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArgSerIleGlyProArgG	360
Db	1071	ATTGGCCTATTATGAGCTATGCTTGGAGAAGAACACACAGATCCATCGCCCCCGTG	1130
QY	360	luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPheAlaIar	380
Db	1131	AGAGTGCACAGAGTACCGACAGCGAGCCTACTCGAAAAAGAACTGCTTTTCTCGGTT	1190
QY	380	YrAspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIleP	400
Db	1191	ATGATGTTTGGGANTTCCCTTTGATCGGATTCATGGACTCCAGAGAGATGATTC	1250
QY	400	roLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG	420
Db	1251	CCAAGAGATTGCAAGAGAAACAGAAAGTCAGAACCGGAAACAGCGCTACGCATCTCC	1310
QY	420	luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL	440
Db	1311	AGATTGGGAACCTTATGAAGAAATCGAGACCTTTGTGAAAGAGAACCAAGAGATCAC	1370
QY	440	euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuT	460
Db	1371	TGGATGTGAGCAGACTGACCGGGGAAGTGGCCCCCTGCTGTATGAAGGCATCAGTCTCA	1430
QY	460	hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer	479
Db	1431	CATGAATCTCAAACTCTCTGAATGGTTC-CCAGCGGGTGGTGGATGGAGCGGCGTAATCTCT	1489
QY	480	AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly	499
Db	1490	GACCAAGAGTGTACAGAGCTGACCAATGTGGCAGCAACCTCAGGAGATGGC	1549
QY	500	TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe	519
Db	1550	TACCGGGGTGAGACCTCCCCACATCTCCCAATGAAAGTTCTATGGTGTCACTGTCTTC	1609
QY	520	LysAlaLeuLysLeuGlyGlnGlyLysValProLeuGlnSerAlaHisLeuTyrTyr	539
Db	1610	AAAGCCCTCAGCTGGGCAAGAGCAAGTCTCTCTGACAGAGTGGCCCTCTACTATAC	1669
QY	540	AsnValThrGluLysValArgIleMetGluSerTyrPheArgLeuAspThrProLeu	559
Db	1670	AAAGTGTGAGGAGAGTGGCGCATCATGGAGTCTACTTCCGCTGGATACGCCCTC	1729
QY	560	TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArg	579
Db	1730	TACTTTTCTACTCTCACTGTGTGCGCACTGCCATCGAAGAGGTCCAGCGAGAGG	1789
QY	580	LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal	599
Db	1790	AAAGATGATAGTATCCAGTCCAGTCCAGTGGACAACTGCATCTCTGAATGCCAGACCTCGTG	1849
QY	600	CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn	619
Db	1850	TGTGTCAAAGAGCGCCCGACCTTACACCTTCCGAGCTACAGCGCATCTTTACTTAAT	1909

```
QY 620 GlyAspPheAspGlyGlyAanPheTyrPheThrGluLeuAspAlaValThrAla 639
Db 1910 GGGGACTTCGATCGCGGAACTTTTATTTCACTGAACCTGGATGCCAAGACCGTGACGGCA 1969
QY 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAanProHis 659
Db 1970 GAGGTGAGCCTCAGTGTGGAAGCGGTGGATTCTCTTCAGGCACCTGAARACCCACAT 2029
QY 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuThrPheThrLeuAsp 679
Db 2030 GGAAGTGAAGCTGTCAACAGGGGCGACGCTGTGCCATCCCTGTGGTTTCAACCTGGAC 2089
QY 680 ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer 699
Db 2090 CCTCGACACAGCGGCGGACAGGTCGAGCAGATGACCTGTGAGATGCTCTTCAGC 2149
QY 700 ProGluGlnMetAspLeuSerGlnGluInProLeuAspAlaGlnGlnGlyProProGlu 719
Db 2150 CCAGAAGAGATGACCTCTCCAGGAGCAGCCCTGGATGCCAGCAGGGCCCCCGGAA 2209
QY 720 ProAlaGlnGluSerLeuSerGlySerGluSerIlySerProLysAspGluLeu 736
Db 2210 CCTGCACAAGAGTCTCTCAGGAGTGAATCGAAGCCCAAGATGAGCTA 2260

RESULT 9
AX877139 2524 bp DNA linear PAT 17-DEC-2003
LOCUS
DEFINITION Sequence 12044 from Patent EP1074617.
ACCESSION AX877139
VERSION AX877139.1 GI:40031875
KEYWORDS
SOURCE
  ORGANISM Homo sapiens (human)
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
  Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
  Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
  Primers for synthesising full-length cDNA and their use
  Patent: EP 1074617-A 12044 07-FEB-2001;
  Research Association for Biotechnology (JP)
FEATURES
  source
    1..2524
    /organism="Homo sapiens"
    /mol_type="unassigned DNA"
    /db_xref="taxon:9606"
    16..2226
    /note="unnamed protein product"
    /protein_start=1
    /db_xref="GI:40031876"
    /translation="MAVRALKLLTLLAVVAASQAEVESEAGWGVTPDLLFAEGTA
    AYARGDWPGVLSMERALRRAALRLRCRTQCAADFPWELDPDWSPPSAQSGAA
    ALRDLSPFGGLRPAACLRCLGPPAAHSISEMELFRKRSYPNYLOVAFKINKLE
    KAVAAATFFVGNPEHEMEXQNDLYCTMSGVKEADFKLETPHMQBELGVLYSE
    EQQEAAPHELAQEFVAYEECALCEGPDYDNGINLYSNADLFQAITDHYIQLV
    NQCNQVTELASHPSEKPFEDFLFPHYNYLQFAYNIGNYTQAVECAKTYLLFFPD
    EWMQNAYLAAMLGEEHRSIGRESAKYRORSELEKSLFFAYDVFQIPVDPDS
    WTPBEVTPKRLQEKQKSERETAVRISQEIGNLMKEIETLVEEKTESLDVSRLTREGG
    PLLYEG:SLTNMSORVMDGVI SDHECOELRLTNVAATSGDVGRTGQTSPT
    PNEKFGVTFKALKIQEGKVPQSAHLYNNTKVRRTMESVFRDLDTPLYPSYHL
    VCRFAIEVQAEKDDSHPHVDNCLNAETLNCVKEPPATPRDYSAILLYLNGDFDG
    GNFTFLDADKVTAEVQPCQRAVGFSSGTENPHGVKATRGQRCAIALWFLDPRH
    SERDRVQADDLVKMLFSPEDMLSQEQMDAQGPPEPAQESLSGSEKPKDEL"

Alignment Scores:
Pred. No.: 0 Length: 2524
Score: 407.00 Matches: 707
Percent Similarity: 99.16% Conservatives: 0
Best Local Similarity: 99.16% Mismatches: 3
Query Match: 55.30% Indels: 6
DB: 6 Gaps: 0
```

Db 1034 ATTTGGCCTATTATGTCAGCTATGCTTGTGAGAGAACACACAGATCCATCGGCCCCCGTG 1093  
 Qy 360 luserAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaT 380  
 Db 1094 AGAGTGCCAGAGGATACCCAGAGGAGCTACTGTGAAAAAGAACTGCTTTTCTTCGCTT 1153  
 Qy 380 yrAspValPheGlyLeuProPheValAspProAspSerTyrThrProGluGluValIleP 400  
 Db 1154 ATGATGTTTTGGAACTCCCTTTGTGATCCGGATTATGACATCCAGAGAAGTGTATC 1213  
 Qy 400 roLysArgGlnGlnLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
 Db 1214 CCAAGAGATTGCAAGAGAAACAGAAAGTCAGAACGGGAACAGCGCTACGATCTCCAGG 1273  
 Qy 420 lulleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440  
 Db 1274 AGATTGGGAACCTTATGAAGGAATCGAGACCCCTTTGTGAAGAGAACCAAGAGATCAC 1333  
 Qy 440 euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeuT 460  
 Db 1334 TGGATGTCAGCAGACTGACCCCGGAAGGTGGCCCTGCTGTATGAAGGCATCAGTCTCA 1393  
 Qy 460 hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer 479  
 Db 1394 CCATGAATCCAAACTCTCTGAATGGTTC-CCAGCGGGTGGTGTGGACGGCGTAATCTCT 1452  
 Qy 480 AspHisGluCysGlnGlnGluGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499  
 Db 1453 GACCACGAGTGTACAGAGCTGCAGAGCTGACCAATGTGGCAGCAACTCAGGAGATGGC 1512  
 Qy 500 TyrArgGlyGlnThrSerProHisThrProAsnGlnLysPheTyrGlyValThrValPhe 519  
 Db 1513 TACCGGGTCAAGCTCCCACTACTCCCAATGAAAGTTCTATGTGTGCTACTGTCTTC 1572  
 Qy 520 LysAlaLeuLysLeuGlyGlnGlnGlyLysValProLeuGlnSerAlaHisLeuTyrTyr 539  
 Db 1573 AAAGCCCTCAAGCTGGGCGAGAGGCAAGTTCTCTGCAGAGTGCCACCTGTACTAC 1632  
 Qy 540 AsnValThrGluLysValArgIleMetGluSerTyrPheArgLeuAspThrProLeu 559  
 Db 1633 AACGTACGGAAGAGTGGCGGATCATGTAGTCTACTTCCGCTGGATAGCCCTC 1692  
 Qy 560 TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArg 579  
 Db 1693 TACTTTTCTACTCTCATCTGTGTGGCTGCGCACTGCCATCGAAGAGTCCAGGCGAGAGG 1752  
 Qy 580 LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599  
 Db 1753 AAGGATGATGATCATCCAGTCCAGTGGACCACTGCATCTGAAATCCGAGACCCCTCGT 1812  
 Qy 600 CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619  
 Db 1813 TGTGTCAAAGAGCCCCAGCCCTACCTTCGCGACTACAGCGCCATCTTTTACCTAAAT 1872  
 Qy 620 GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla 639  
 Db 1873 GGGGACITCGATGGCGGAACTTTTATTTCACTGAATGCCAAGACCGTGACGGCA 1932  
 Qy 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659  
 Db 1933 GAGGTGACGCTCAGTGTGGAAGAGCCGTGGGATTTCTTTTCAAGCACTGAAACCCACAT 1992  
 Qy 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAsp 679  
 Db 1993 GGAGTGAAGGCTGTACACAGGGGCGAGCGTGTGCCATCGCCCTGTGGTTTCACTTGGAC 2052  
 Qy 680 ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer 699  
 Db 2053 CCTCGACACAGCGGAGGAGGAGGTCAGGATGATGATGATGATGATGATGATGATGATG 2112  
 Qy 700 ProGluGluMetAspLeuSerGlnGlnPro 710  
 Db 2113 CCAGAGAGATGGACCTTCTCCAGGAGCAGGCC 2145

## RESULT 10

BD156505  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

## REFERENCE

1 (bases 1 to 2524)

## AUTHORS

Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,

## TITLE

Primer for synthesizing full-length cDNA and use thereof.

## JOURNAL

Patent: JP 2002191363-A 11348 09-JUL-2002;

## COMMENT

OS Homo sapiens (human)  
 PN JP 2002191363-A/11348  
 PD 09-JUL-2002  
 PF 28-JUL-2000 JP 2000280990  
 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU  
 PI SAITO,  
 PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
 PI KEIICHI NAGAI, TETSUJI OTSUKI  
 PC

## FEATURES

source

## ORIGIN

Alignment Scores:

Pred. No.: 0

Score: 407.00

Percent Similarity: 99.18%

Best Local Similarity: 99.18%

Query Match: 55.30%

DB: 6

US-10-045-815-4 (1-736) x BD156505 (1-2524)

Qy 1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaSer 20

Db 16 ATGGCGGTACCGCGCTTGAAGCTGCTGACACACATGCTGCTGCTGCTGCTGCTGCTGCT 75

Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuPheAla 40

Db 76 CAAGCCGAGGTGAGTCCGAGCAGATGGGCGATGGTGGCTGATCTCTCTTCTGCTC 135

Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60

Db 136 GAGGGGAGCGCAGCTACGCGCGGGGACTGGCCCGGGGTGGTCTTGAGCATGGAACGG 195

Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80

Db 196 GCGCTGCGCTCCCGGCGAGCCCTCCGCGCTTCCGCTGCGCTGCGCTGCGCTGCGCTG 255

Qy 81 AlaAspPheProTyrGluLeuAspProAspTyrPsrProSerProAlaGlnAlaSerGly 100

Db 256 GCGGACTTCCCGTGGGAGCTGGAGCCCGACCTGGTCCCGACCGCCCGCCCGCCCGCCG 315

Qy 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLe 120

Db 316 GCGGCG-CGCGCTGCGGACCTGAGCTTCTTCGGGGGCTTCTGCGTGGCTGCTGCTGCT 374

BD156505 2524 bp DNA linear PAT 17-JAN-2003  
 Primer for synthesizing full-length cDNA and use thereof.  
 BD156505  
 BD156505.1 GI:27862263  
 JP 2002191363-A/11348.  
 Homo sapiens (human)

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,  
 Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.  
 Primer for synthesizing full-length cDNA and use thereof.  
 Patent: JP 2002191363-A 11348 09-JUL-2002;

HELIX RESEARCH INSTITUTE  
 OS Homo sapiens (human)  
 PN JP 2002191363-A/11348  
 PD 09-JUL-2002  
 PF 28-JUL-2000 JP 2000280990  
 PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU  
 PI SAITO,  
 PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
 PI KEIICHI NAGAI, TETSUJI OTSUKI  
 PC

C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ 10,  
 C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N5/00, C12N5/00 CC  
 Primer for synthesizing full-length cDNA and use thereof FH Key

FT CDS  
 Location/Qualifiers  
 (16)..(2223).

Location/Qualifiers  
 1..2524  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

Alignment Scores:  
 Pred. No.: 0 Length: 2524  
 Score: 407.00 Matches: 707  
 Percent Similarity: 99.18% Conservative: 0  
 Best Local Similarity: 99.18% Mismatches: 3  
 Query Match: 55.30% Indels: 6  
 DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x BD156505 (1-2524)

Qy 1 MetAlaValArgAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaSer 20  
 Db 16 ATGGCGGTACCGCGCTTGAAGCTGCTGACACACATGCTGCTGCTGCTGCTGCTGCTGCT 75

Qy 21 GlnAlaGluValGluSerGluAlaGlyTyrGlyMetValThrProAspLeuPheAla 40  
 Db 76 CAAGCCGAGGTGAGTCCGAGCAGATGGGCGATGGTGGCTGATCTCTCTTCTGCTC 135

Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTyrProGlyValValLeuSerMetGluArg 60  
 Db 136 GAGGGGAGCGCAGCTACGCGCGGGGACTGGCCCGGGGTGGTCTTGAGCATGGAACGG 195

Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
 Db 196 GCGCTGCGCTCCCGGCGAGCCCTCCGCGCTTCCGCTGCGCTGCGCTGCGCTGCGCTG 255

Qy 81 AlaAspPheProTyrGluLeuAspProAspTyrPsrProSerProAlaGlnAlaSerGly 100  
 Db 256 GCGGACTTCCCGTGGGAGCTGGAGCCCGACCTGGTCCCGACCGCCCGCCCGCCCGCCG 315

Qy 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLe 120  
 Db 316 GCGGCG-CGCGCTGCGGACCTGAGCTTCTTCGGGGGCTTCTGCGTGGCTGCTGCTGCT 374

QY 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140  
Db 375 GCCTGCTGCTCGGCGCGCGCGCCACTCGCTCAGCGAAGATGAGCTGGAGTT 434  
QY 140 eArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuG 160  
Db 435 CCAGCAGCGAGCGCCCTTACAACTACCTGCGAGTGGCTACTTCAAGATCAACAAAGTTGGA 494  
QY 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetG 180  
Db 495 GAAAGCTGTGCTGCAGCACACACCTTCTTCGTGGCGAATCTTGAGCACATGGAATGCA 554  
QY 180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
Db 555 GCAGAACCTAGACTATTACCAACCACTGCTGAGTCAAGAGCGGCACTTCAAGGATCT 614  
QY 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluG 220  
Db 615 TGAGACTCAACCCCATATGCAAGATTTTCAGCTGGGAGTGGGACTTACTCAGAGGAACA 674  
QY 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrG 240  
Db 675 GCCACAGAAAGCTGTGCCACCTAGAGCGCGCTGCAAGAAATACTTTGTGGCTTATGA 734  
QY 240 uGluCysArgAlaLeuCysGluGlyProTyrAsnTyrAspGlyTyrAsnTyrLeuGluTy 260  
Db 735 GAGTGGCGGTGCTTCGCGAAGGCGCCATGACTAGCTAGCTGCTCAACTACCTTGA 794  
QY 260 rAsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCysLysG 280  
Db 795 CAACGCTGACCTTCCAGCGCCATCAGATCATTAATCACTCCAGTCTCAACTGTAAGCA 854  
QY 280 nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
Db 855 GAACTGTGTCAGGAGCTTCCCTCCACCAAGTCGAGAGAACCCCTTTGAAGACTTCCT 914  
QY 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320  
Db 915 CCATCGCATTAATATCTGCAGTTTGCCTACTATAACATTTGGGAATTATACACAGGC 974  
QY 320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
Db 975 TGT-TGAATGTGCAAGACCTATCTTCTTCTTCCCAATGACGAGTGTATGAACAAA 1033  
QY 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGluHisThrArgSerIleGlyProArgG 360  
Db 1034 ATTTGGCTATTATGCACTATGCTTGGAGAAGAACACACCATCATCCGCGCCCGTG 1093  
QY 360 luserAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlar 380  
Db 1094 AGAGTGCAGAGGATCCGACAGCGAAGCTACTTGGAAAAAGAACTGCTTTTCTCGCTT 1153  
QY 380 yrAspValPheGlyIleProPheValAspProAspSerTyrThrProGluGluValIleP 400  
Db 1154 ATGATGTTTTTGGAAATTCCTTTGTGATCCGATTCATGAGCTCCAGAAAGATGATTC 1213  
QY 400 rLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGlnG 420  
Db 1214 CCAGAGATTGCAAGAGAAACAGAGTCAAGCGGAAACAGCGGTACGATCTCCACAG 1273  
QY 420 luIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSerL 440  
Db 1274 AGATTGGGAACCTTATGAAGAAATCGAGACCCCTTGTGGAGAGAGAACCAAGAGTCA 1333  
QY 440 euAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrGluGlyIleSerLeu 460  
Db 1334 TGGATGTGAGCAGACTGACCCCGAAGAGTGGCCCTTCTGTATGAAGGATCAGTCTCA 1393  
QY 460 hrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSer 479  
Db 1394 CCATGAACCTCAAACTCTGAATGGTTC-CCAGCGGGTGTGTGAGCGCGTAATCTCT 1452

QY 480 AspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGly 499  
Db 1453 GACCACGAGTGTGAGGAGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGGC 1512  
QY 500 TyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPhe 519  
Db 1513 TACCGGGTCAAGCTCCCACTATCCCAATGAAAGTTCTATGGTGTCTCAGTCTCTTC 1572  
QY 520 LysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTyr 539  
Db 1573 AAAGCCCTCAAGCTGGGCAAGGCAAGTTCCTCTGCAGAGTGCCACCTGTACTATAC 1632  
QY 540 AsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLeu 559  
Db 1633 AACGTGACGGAAGAGTGGCGGCATCATGAGTCTTACTTCCGCTCGATACGCCCTC 1692  
QY 560 TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluArg 579  
Db 1693 TACTTTTCT 1752  
QY 580 LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599  
Db 1753 AAGGATGATAGTCACTCCAGTCCAGTGCAGCAACTCCATCCCTGAAATGCCGAGACCTCTG 1812  
QY 600 CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619  
Db 1813 TGTGTCAAGAGCCCTCCAGCTACACCTTCCGCGACTACAGCGCATCTCTTTACCTAAAT 1872  
QY 620 GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAla 639  
Db 1873 GGGGACTTCCGATGGCGGAAACTTTTATTTCATGTAACCTGGATGCCAAGACCTGTGAGCA 1932  
QY 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659  
Db 1933 GAGGTGACGCTCAGTGTGGAAGACCGTGGGATCTCTTCAGGCACCTGAAACCCACAT 1992  
QY 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAsp 679  
Db 1993 GGAGTCAAGGCTGTCAACAGGGGGCAGGCTGTGCATCGCCCTGTGGTTTACCCTCGAC 2052  
QY 680 ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer 699  
Db 2053 CTTCCAGCAGCGAGCGGACAGGCTGCGGAGATGACCTGGTGAGATGCTCTCTCTTCCAG 2112  
QY 700 ProGluGluMetAspLeuSerGlnGluGlnPro 710  
Db 2113 CCAGAAGAGATGAGCTCTCCAGCAGCAGCCC 2145  
RESULT 11  
AK027680  
LOCUS  
DEFINITION  
Homo sapiens cDNA FLJ14774 fis, clone NT2RP4000051, weakly similar  
to SYNAPTONEURAL COMPLEX PROTEIN SC65.  
ACCESSION  
AK027680  
VERSION  
AK027680.1 GI:14042539  
KEYWORDS  
oligo capping; fis (full insert sequence).  
SOURCE  
Homo sapiens  
ORGANISM  
Homo sapiens  
REFERENCE  
1  
Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,  
Matsumura H., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
Wakayama A., Nakamura Y., Nagahara K., Masuho Y. and Sasaki N.  
NEDO human cDNA sequencing project  
2 (bases 1 to 2524)  
Isogai T. and Otsuki T.  
Direct Submission  
Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)



COMMENT

NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5' - 3' - end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

## FEATURES

Location/Qualifiers

```

1. 2224
/organisms="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP4000051"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_1b="NT2RP4"
/notes="Cloning vector: pM18
precursor cells after 2-week
induction."
16. 2226

```

CDS

## ORIGIN

Alignment Scores:

Pred. No.:	0	Length:	2524
Score:	407.00	Matches:	707
Percent Similarity:	99.16%	Conservative:	0
Best Local Similarity:	99.16%	Mismatches:	3
Query Match:	55.30%	Indels:	6
DB:	9	Gaps:	0

US-10-045-815-4 (1-736) x AK027680 (1-2524)

1	MetAlaValA:GAlaLeuLysLeuThrThrLeuLeuAlaValAlaAlaAlaSer	20
16	ATGGCGGTACGCGGCTTTGAAGCTCTGACCACACTGTGTGGCTCTCGTGGCGGTGCTGCC	75
21	GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla	40
76	CAAGCCGAGGTCTGAGTCCGAGGCAAGATGGGGCATGGTGACGCTGATCTGTCTTCGCG	135
41	GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg	60
136	GAGGGGACCGAGGCTACGGCGCGGGAGATGGCCGGGTGCTCTGACATGGGAAGCG	195
61	AlaLeuArgSerA:GAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla	80
196	CGCTGCGCTCCCGGGCAGCCCTCCGCGCCTTCGCTGCGTCCGACCCAGGTGTC	255
81	AlaAspPheProTrpGluLeuAspProAspTrpSerProSerProAlaGlnAlaSerGly	100
256	GCCGACTTCCCGTGGGAGCTGACCCCGACTGTGTCCCCAGCCCGCCAGCGCTCGGGC	315
101	AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLe	120
316	GCCGC-CGCCCTCGGGACCTGAGCTTCTTCGGGGGCGCTTCTCGCTCGGGCGCTGCT	374



```
Db 1453 GACACGAGTGTGAGGAGTGCAGAGACTGACCAATGTGCGCAACCTCAGAGATGCG 1512
Qy 500 TTAAGGAGTGTGAGGAGTGCAGAGACTGACCAATGTGCGCAACCTCAGAGATGCG 1512
Db 1513 TACCGGGTTCAGACCTCCACATATCTCCATGAAAGTTCTATGGTGTCTACTGCTTC 1572
Qy 520 LysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGluSerAlaHisLeuTyr 539
Db 1573 AAAGCCCTCAAGCTGGCGGCAAGAAGCAAAAGTTCTCTGCGAGAGTCCCACTGTACTAC 1632
Qy 540 AsnValThrGluLysValArgArgLysMetGluSerTyrPheArgLeuAspThrProLeu 559
Db 1633 AAGCTGACGAGAGAGTGGCGGCATCATGAGTCTCTCTCCGCTGATAGCCCTTC 1692
Qy 560 TyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluValGlnAlaGluArg 579
Db 1693 TACTTTTCTTACTCTCATCTGTTGTCGCGCACTGCCATGCGAGAGTCCAGGCAGAGAG 1752
Qy 580 LysAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVal 599
Db 1753 AAGGATGATAGTATCATCCAGTCCACGTGGCAACTGCATCTGAATGCCGAGACCTCGTG 1812
Qy 600 CysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAsn 619
Db 1813 TGTGTCAAAGAGCCCCAGAGCTACACCTTCGCGGACTACAGCGCCATCTTTACCTAAT 1872
Qy 620 GlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThra 639
Db 1873 GGGGACTTCGATGCGGAACTTTTATTTCACTGAACGTGATGCCAAGACCGTGAGCGCA 1932
Qy 640 GluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHis 659
Db 1933 GAGGTGAGGCTGATCACCAGGGGCGAGCGTGTGCCATCGCTGTGTTTACCTGGAC 2052
Qy 660 GlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAsp 679
Db 1993 GGAGTGAGGCTGATCACCAGGGGCGAGCGTGTGCCATCGCTGTGTTTACCTGGAC 2052
Qy 680 ProArgHisSerGluArgAspArgValGlnAlaAspAspLeuValLysMetLeuPheSer 699
Db 2053 CCTGCACACAGCAGCGGACAGGAGTGCAGGACATGACCTGTGAGATGCTCTTCAGC 2112
Qy 700 ProGluGluMetAspLeuSerGlnGluGlnPro 710
Db 2113 CCAGAGAGATGAGCTCTCCAGGAGCAGCCC 2145

RESULT 12
AF097431 2829 bp mRNA linear PRI 22-NOV-2000
LOCUS Homo sapiens GROS1-S protein mRNA, complete cds.
ACCESSION AF097431
VERSION AF097431.1 GI:11127635
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
1 (bases 1 to 2829)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Kaul, S.C., Sugihara, T., Yoshida, A., Nomura, H. and Wadhwa, R.
Gros1, a potential growth suppressor on chromosome 1: its identity
to basement membrane-associated proteoglycan, leprecan
Oncogene 19 (32), 3576-3583 (2000)
JOURNAL
MEDLINE 20406537
PUBMED 10951563
REFERENCE
2 (bases 1 to 2829)
Wadhwa, R., Sugihara, T. and Kaul, S.
Direct Submission
TITLE
Submitted (06-OCT-1998) Group 4, Chugai Research Institute of
JOURNAL Molecular medicine, 153-2 Nagai, Niihari, Ibaraki 300-41, Japan
FEATURES
Location/Qualifiers
1..2829
/organism="Homo sapiens"
```

```
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="1"
/tissue_type="testis"
52..1143
/notes="member of protein B family"
/codon_start=1
/product="GROS1-S protein"
/protein_id="AAG31018.1"
/db_xref="GI:11127635"
/translation="MAVRALKLLTLLAVAAASQAESEAGWMTPTDLLFAEGTA
AYAGDPGVVLSMERALRSALRLRCRTOCAADFPWELDPDWSPPSAQASGAG
ALRDLSPFGILLRRAACLRCLGPPAAHSEMELEFRKSPNYLQVAFKINKLE
KAVAAATFFVGNPEHMQNLDYQDTMSGKEADFKOLETQPHMFEFLGVRLYSE
EQGAVFHLARALCEYFVAYEERACLEGPDYDGNLYEYNADLFCALIDHYLOVL
NCKQNVTLASHFSRKPFDPLFSPHNYNLYQFAYINIGNITQAGECAKTYLLFFND
EVMNQNLAYAAMLGEBEHSIGPREQGT"

ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 2829
Score: 360.00 Matches: 360
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 48.91% Indels: 0
DB: 9 Gaps: 0

US-10-045-815-4 (1-736) x AF097431 (1-2829)
Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaLaser 20
Db 52 ATGGCGGTACGGCGCTTGAAGCTGTGACCACTGTGCTGGCTGTCTGCGCGCTGCCCTCC 111
Qy 21 GlnAlaGluValGluSerGluAlaGlyTTPGlyMetValThrProAspLeuLeuPheAla 40
Db 112 CAAGCGAGGTGAGTCCGAGCAGATGGGCGATGGTGGCGCTGATCTCTCTCGCC 171
Qy 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTTPProGlyValValLeuSerMetGluArg 60
Db 172 GAGGGGACCGGAGCTTACGCGCGCGGGGACTGGCCCGGGTGGTCTCTGAGCATGAACGG 231
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
Db 232 GCGCTCGCTCCCGGCGAGCCCTCCGCGCTTCCTGCGCTGCCGACCCAGGTGTGCC 291
Qy 81 AlaAspPheProTTPGluLeuAspProAspTTPSerProSerProAlaGlnAlaSerGly 100
Db 292 GCGGATTTCCGTTGGAGCTGAGCCCGGCTGCTCCCGCGCGCGCGCGCGCTCGGCG 351
Qy 101 AlaGlyAlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgArgAlaAlaCysLeu 120
Db 352 GCGCGCGCGCTCGCGGACCTGAGCTTCTTGGGGGCGCTTCTGCGTGGCGCTGCCCTG 411
Qy 121 ArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluMetGluLeuGluPhe 140
Db 412 CCGCGCTGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 471
Qy 141 ArgLysArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGlu 160
Db 472 CGCAAGCGGAGCGCGCTACAACTACCTGCGAGTTCGCTTCTTCAAGATCAACAAGTTGG 531
Qy 161 LysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGln 180
Db 532 AAAGCTGTTGCTGCGAGCACACACCTTCTTCTGCGGCAATCTCCGAGCACATGAATGCG 591
Qy 181 GlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLeu 200
Db 592 CAGAAGCTAGACTATACCAACCATGCTGAGTGAAGAGCGCGCGCGCGCGCGCGCG 651
Qy 201 GluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGln 220
Db 652 GAGACTCAACCCCATATGACAGATTTTCGACTGGGAGTGGGACTCTTCTACTCAGAGAAC 711
```

QY 221 ProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyrPheValAlaTyrGlu 240  
 DB 712 CCACAGGAAGCTGTGCCCCACCTAGAGCGCGCTGCAAGAATACTTTGTGGCCTATGAG 771  
 QY 241 GluCyAsnGlnAlaLeuCyGlnGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
 DB 772 GAGTCCCGTGGCCCTCTGGAAGGGCCCTATGACTACGATGCTCAACTACCTTTGAGTAC 831  
 QY 261 AsnAlaAspLeuPheGlnAlaIleThrAspHisTyrIleGlnValLeuAsnCySLysGln 280  
 DB 832 AACGCTGACCTCTTCCAGGCCATCACACATCATACATCCAGGCTCTCACTGTAAAGCAG 891  
 QY 281 AsnCyValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLeu 300  
 DB 892 AACTGTGTACGGAGCTTCTCCACCACCAAGTGCAGAGAGCCCTTTGAGACTTCTCCTC 951  
 QY 301 ProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAla 320  
 DB 952 CCATCGCATTAATATCTCGAGTTTCCCTACTATACATTTGGAAATATACAAAGCT 1011  
 QY 321 GlyGluCyAlaLysThrTyrLeuPhePheProAsnAspGluValMetAsnGlnAsn 340  
 DB 1012 GGTGAATGTGCAGACCTATCTCTCTTCCCAATGACGAGGTGATGAACCAAAAT 1071  
 QY 341 LeuAlaTyrTyrAlaAlaMetLeuGluGluHisThrArgSerIleGlyProArgGlu 360  
 DB 1072 TTGGCTAATTATGAGCTATGCTTGGAGAGAACACACAGATCATCGGCCCCCGTGAG 1131  
 RESULT 13  
 LOCUS BC015309 1562 bp mRNA linear PRI 30-SEP-2003  
 DEFINITION Homo sapiens leucine proline-enriched proteoglycan (leprecan) 1,  
 mRNA (CDNA clone IMAGE:4395073), partial cds.  
 ACCESSION BC015309  
 VERSION BC015309.2 GI:34782974  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 1562)  
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
 Diatchenko, L., Narusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
 Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
 Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,  
 Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E.,  
 Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 JOURNAL MEDLINE 22388257  
 PUBMED 12477932  
 REFERENCE 2 (bases 1 to 1562)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-OCT-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 COMMENT On Sep 16, 2003 this sequence version replaced gi:15929775.

Contact: MGC help desk  
 Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
 Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www-shgc.stanford.edu>  
 Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAC Place: 27 Row: C Column: 1.

FEATURES  
 source  
 1..1562  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4395073"  
 /tissue\_type="Duodenum, adenocarcinoma"  
 /clone\_lib="NIH MGC\_88"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"  
 <1..1562  
 /gene="LEPRE1"  
 /note="synonym: GROSI"  
 /db\_xref="LocusID:64175"  
 <1..1226  
 /gene="LEPRE1"  
 /codon\_start=3  
 /product="LEPRE1 protein"  
 /protein\_id="AAH15309.2"  
 /db\_xref="GI:34782975"  
 /db\_xref="LocusID:64175"  
 /translation="DPRVREVNQNLAYAAVLGSEHTSITGPRESAKEYRQSRLLK  
 ELLPAYDVFGIPVDPDSWTFEVIKRLQEKOKSRETVARISQIGNLMKEIETL  
 VEETKESLDSRLITREGLPYEGISITMNSKLNGSQRVMDGVISDHCEQLQRL  
 INVAATSGDYGQSTPHTNEKFGYTVFKALKGQGVLPQSAHLIYNYVTEKVR  
 IMESYFRLDTPLFYSLVCLVCTAIEEVAERKDDSHPVNDNCILNAETLVCYKEPP  
 AYTFYDGAIIYLANGDFDGGNFYFTELDKATVTAEOVQPCGAVGFSSTENPHGVA  
 TQRGRCAALWFTLDPRHSRDRVQADLLVKMLFSPEMDLSQEQPLDAQQGPPEA  
 QESLSGSSKPKDEL"  
 426..1046  
 /gene="LEPRE1"  
 /note="P4HC; Region: Polyl 4-hydroxylase alpha subunit  
 homologues. Mammalian enzymes catalyse hydroxylation of  
 collagen, for example. Prokaryotic enzymes might catalyse  
 hydroxylation of antibiotic peptides. These are  
 2-oxoglutarate-dependent dioxygenases, requiring  
 2-oxoglutarate and dioxygen as cosubstrates and ferrous  
 iron as a cofactor"  
 /db\_xref="CDD:smart00702"  
 misc\_feature

## ORIGIN

## Alignment Scores:

Pred. No.: 0 Length: 1562  
 Score: 301.00 Matches: 401  
 Percent Similarity: 99.50% Conservat: 0  
 Best Local Similarity: 99.50% Mismatches: 1  
 Query Match: 40.90% Indels: 2  
 DB: 9 Gaps: 0

US-10-045-815-4 (1-736) x BC015309 (1-1562)

QY 335 GluValMetAsnGlnAsnLeuAlaTyrTyrAlaAlaMetLeuGlyGluGluHisThrArg 354  
 DB 18 GAGGTGATGACCAACAAATTTGGCCTATTATGACCTATGCTTGGAGAGAACACACCAGA 77  
 QY 355 SerIleGlyProArgGluSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGlu 374  
 DB 78 TCCATCGGCCCCCGTGAGAGTGCCCAAGAGTACCGACAGCGAAGCCTACTGGAAGAA 137

```
QY 375 LeuLeuPhePheAlaTyrAspValPheGlyIleProPheValAspProAspSerTyrThr 394
Db 138 CTGCTTTCTCGCTTATGATGTTTGGAAATCCCTTTGTGGATCCGGAATCATGCACT 197

QY 395 ProGluValIleProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAla 414
Db 198 CCAGAGAAGTGTATCCCAAGATGTCAGAGAAACAGAAAGTCAGAACGGGAAACAGCC 257

QY 415 ValArgIleSerGlnGluIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGlu 434
Db 258 GTACGCATCTCCAGGAGATGGGAACCTTATGAGGAAATCAGACCTTGTGGAAGAG 317

QY 435 LysThrLysGluSerLeuAspValSerArgLeuThrArgGluGlyProLeuLeuTyr 454
Db 318 AAGACCAAGAGTCACTGGATGTGAGCAGACTACCCGGGAAGTGGCCCTCGTGTAT 377

QY 455 GluGlyIleSerLeuThrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMe 474
Db 378 GAGGCATCAGTCTCACCAGAACTCCAACTCTGATGGTTC-CCAGCGGTGTGTAT 436

QY 474 tAspGlyValIleSerAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAla 494
Db 437 GGACGGCGTAATCTGACCACGAGTGTGAGGAGCTGCAGAGACTGACCAATGTGCAGC 496

QY 494 aThrSerGlyAspGlyTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTy 514
Db 497 AACTCAGGAGATGGCTACCGGGTTCAGACTCCCACTACTCCCAATGAAGAAAGTCTTA 556

QY 514 rGlyValThrValPheLysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSe 534
Db 557 TGGTGTCACCTGCTTCAAGCCCTCAAGCTGGGCAAGGCAAAAGTTCCTCTGCAGAG 616

QY 534 rAlaHisLeuTyrTyrAsnValThrGluLysValArgArgIleMetGluSerTyrPhear 554
Db 617 TGCCCACTCTACTACAACGTGACGGAAGAGTGGCGGCATCATGGAGTCTTACTCCG 676

QY 554 gLeuAspThrProLeuTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGlu 574
Db 577 CTGGATACGCCCTCTACTTTTCTACTCTCATCTGGTGTGCGGCACTGCCATCGAGA 736

QY 574 uValGlnAlaGluArgLysAspSerHisProValHisValAspAsnCysIleLeuAs 594
Db 737 GGTCCAGGACAGAGGAAGATGATAGTATCATCCAGTCCAGTGGCAAACTGCATCTCGAA 796

QY 594 nAlaGluThrLeuValCysValLysGluProProAlaTyrPheArgAspTyrSerAl 614
Db 797 TGGCGAGACCTCGTGTGTCAAAGAGCCCGGAGCTACACCTTCGCGACTACAGCGC 856

QY 614 aIleLeuTyrLeuAsnGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAl 634
Db 857 CATCTTTACTTAATCGGACCTTCGATGGCGGAAACTTTTATTTCACTGAATGGATGC 916

QY 634 aLysThrValThrAlaGluValGlnProGlnCysGlyArgAlaValGlyPheSerSerG 654
Db 917 CAAGACCGTACGCGACAGGTGAGCCCTCAGTGTGGAAGAGCGGTGCGGATTCCTTCAG 976

QY 654 YThrGluAsnProHisGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAla 674
Db 977 CACTGAAAACCCATCGAGTGAAGGTGTCCAGAGGGGACGGCTGTGCCATCGCCCT 1036

QY 674 uTTPPheThrLeuAspProArgHisSerGluArgAspArgValGlnAlaAspAspLeuVa 694
Db 1037 GTGGTTCCACCTCGACCTCGACACAGCGAGCGGACAGGGTGCAGGACATGACCTGGT 1096

QY 694 lLysMetLeuPheSerProGluGluMetAspLeuSerGlnGluGlnProLeuAspAlaGl 714
Db 1097 GAAGATGCTCTTACGCCCAAGAGATGGACCTCTCCAGAGAGAGCCCTCGATGCCCA 1156

QY 714 nGlnGlyProProGluProAlaGlnGluSerLeuSerGlySerGluSerLysProLysAs 734
Db 1157 GCAGGGCCCCCGAACCTGACAAAGAGTCTCTCTCAGGAGTGAATCGAAGGCCAAGGA 1216
```

```
QY 734 pGluLeu 736
Db 1217 TGAGCTA 1223

RESULT 14
AX880254
LOCUS AX880254
DEFINITION Sequence 15159 from Patent EP1074617.
ACCESSION AX880254
VERSION AX880254.1 GI:40034990
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE Primers for synthesising full-length cDNA and their use
JOURNAL Patent: EP 1074617-A 15159 07-FEB-2001;
RESEARCH Association for Biotechnology (JP)
FEATURES
source
1. .2993
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
42..2456
/note="unnamed protein product"
/codon_start=1
/protein_id="CAE90869.1"
/db_xref="GI:40034991"
/translation="MAVRALKLTLTLLAVAAASAEVESEAGMGWTPDILLFAEGTA
AVRGDWPVVLSEMERALRSALRALRLRCTCAADFPWELDPDMSVSPQAASGAA
ALRDLSPFGLLRRAACLRCLGPPAAHLSSEMELEPERKSPYNLYQVAYEKINKLE
KAVAAHTFFVGNPEHMEMQNDLYTQMSGVKEADFKDLTQPHMQFRGLGRLYSE
EQPOAHPLEALQOEYVAYEACALCEGYDYDYNLEYNADLFQALTDHYIQVL
NCKONCVELASHPSREKPFDFLPSHYNLYQFAYNIGNYTQACAKTYLLFFPND
EVMONLAYAAMLGCEHRSIGPESAKYRQSLLEKELFFAYDFVGFIPVDPDS
WTPGEVIKRLQEKQKSERETAVRISOBIGNLKEIETLVEEKESLDVSRKREGG
PLLVEISLTWNSKLLNGSORVMDVTSDEHCEQELORLTNVAATSGDVGRTQSPHT
ENRPFYGVTFKALGCGEKPQSAHLYNVTKEVRIMESFRDLTPLFYFSYSHL
VCRTAIEVQAERKDDSPHVNDNCILNAETFLVCVKGPFATFYDYSAILLYNGDFDG
NCFYTELDKATVTAEVQPCQGRAVGFSGTENPHGVKAVTRGQRCALAFNTLDPRH
SERVRAARAGESSWCQDFFRPWFAPLFPKSHCOMLHERSTWDTSSNALSWSHC
LVLFPAPVNGIQVGKEVKTGSDAEFLVPSLPTSAVLQRVGPAKGEMSLGLRLNLP
PLGSSS"

ORIGIN
Alignment Scores:
Pred. No.: 2,556-314 Length: 2993
Score: 295.00 Matches: 681
Percent Similarity: 98.84% Conservative: 0
Best Local Similarity: 98.84% Mismatches: 4
Query Match: 40.08% Indels: 8
DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x AX880254 (1-2993)
QY 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaAlaSer 20
Db 42 ATGGCGGTACCGCGCTTGAAGCTGCTGACACACTGCTGGCTGCTGCGCGCTGCCCTC 101
QY 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40
Db 102 CAAGCGGAGGTGAGTCCGAGCAGGATGGGCGCATGTTGACGCGCTGATCTGCTTCGCC 161
QY 41 GluGlyThrAlaAlaTyrAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60
Db 162 GAGGGGACCGCAGCCTACGCGCGGGGAGACTGGCCCGGGGTGGTCTCTGAGCATGGAACGG 221
QY 61 AlaLeuA-GserArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80
Db 222 GCGCTGCGCTCCCGGGGAGCCCTTCGCGCCCTTCGCTGCGCTGCGCGCACCCAGTGTGCC 281
```

QY 81 AlaAspPheProTgPLeuLeuAspProAspTyrSerProSerProAlaGlnAlaSerGly 100  
DB 282 GCCGACTTCCTCGTGGAGCTGGACCCCGAGCTGGTCCCCCGAGCCCGCCAGGCTCGGGC 341  
QY 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLe 120  
DB 342 GCCGC-CGCCCTCGCGAGCTGAGCTTCTTCGGGGGCTTCCTCGCTCGCGCTGCTGCCT 400  
QY 120 uArgArgCysLeuGlyProProAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140  
DB 401 CGCGCTGCTCGCGCGCGCGCGCGCGCTTCCTCAGCGAGAGATGAGCTGGAGTT 460  
QY 140 eArgGlyArgSerProTyrAsnTyrLeuGlnValAlaTyrPheLysIleAsnLysLeuGl 160  
DB 461 CCGCAAGCGGAGCCCTACAACTACCTGCGAGTCCGCTACTTCAAGATCAACAAGTTGGA 520  
QY 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetGl 180  
DB 521 GAAAGCTGTTGCTGAGACACACCTTCTTCGTGGGCAATCCTGAGCACATGGAATGCA 580  
QY 180 nGlnAsnLeuAspTyrTyrGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
DB 581 GCAGAACCTAGACTATTACCAAAACCATGCTCTGGAGTCAAGAGCGCGACTTCAAGGATCT 640  
QY 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyrSerGluGluGl 220  
DB 641 TGAGACTCAACCCCATATGCAAGAAATTCGAGTGGGAGTGGACTTACTCAGAGGAACA 700  
QY 220 nProGlnGluAlaValProHisLeuGluAlaLeuGlnGluTyrPheValAlaTyrGl 240  
DB 701 CCCACGGAAGCTGTGCCACCTAGAGCGCGCTGCAAGATACATTTTGTGGCTATGA 760  
QY 240 uGluCysArgAlaLeuCysGluGlyProTyrAspTyrAspGlyTyrAsnTyrLeuGluTyr 260  
DB 761 GGAGTGGCGTCCCTCTCGGAAGGGCGCTATGACTAGCATGGCTTACAACTACCTTGAGTA 820  
QY 260 rAsnAlaAspLeuPheGlnAlaThrAspHisTyrIleGlnValLeuAsnCysLysGl 280  
DB 821 CAACGCTGACCTTTCAGGCGCATCAGATCATATCATCATGCTCCATCTCAACTGTAGCA 880  
QY 280 nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
DB 881 GAACTGTGTACGAGCTGCTCCCAACCAAGTCGAGAGAGCCCTTTTGAAGACTTCCT 940  
QY 300 uProSerHisTyrAsnTyrLeuGlnPheAlaTyrTyrAsnIleGlyAsnTyrThrGlnAl 320  
DB 941 CCCATCGCATTAATATATCTGCAGTTTGGCTACTATAAATTGGGAATTTATACAGGC 1000  
QY 320 aGly-GluCysAlaLysThrTyrLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
DB 1001 TGT-TGAATGTGCCAAGACCTATCTTCTCTTCTTCTCCCAATGACGAGGTGATGAACCAA 1059  
QY 340 snLeuAlaTyrTyrAlaAlaMetLeuGlyGlnGluHisThrArgSerIleGlyProArgG 360  
DB 1060 ATTTGGCTCTATTATGCACTATGCTTGGAGAAGAACACACAGCATCATCCGCCCGCTG 1119  
QY 360 luSerAlaLysGluTyrArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlar 380  
DB 1120 AGATGCCAAGAGTACCGACAGAGAGCTTACTGGAAAAGAACTGCTTTCTTCGCTT 1179  
QY 380 yrAspValPheGlyIlePhePheValAspProAspSerTyrThrProGlu-GluValIle 399  
DB 1180 ATGATGTTTTTGGAAATTCCTTTTGTGATCCGATTCATGAGCTCCAGG-AGAAGTGATT 1238  
QY 400 ProLysArgLeuGlnGluLysGlnLysSerGluArgGluThrAlaValArgIleSerGln 419  
DB 1239 CCCAAGAGATTGCAAGAGAACAGAACTAGACCGGAAACAGCGGTACGCATCTCCAG 1298  
QY 420 GluIleGlyAsnLeuMetLysGluIleGluThrLeuValGluGluLysThrLysGluSer 439  
DB 1299 CAGATTGGGAACCTTATGAAGGAATTCAGACCCCTTGTGGAAGAGAAGAACCAAGGAGTCA 1358

QY 440 LeuAspValSerArgLeuThrArgGluGlyProLeuLeuTyrGluGlyIleSerLeu 459  
DB 1359 CTGGATGTGAGCAGACTGACCCGGGAAGTGGCCCCCTGCTGTATGAAGGCATCAGTCTC 1418  
QY 460 ThrMetAsnSerLysLeuLeuAsnGlyTyr-GlnArgValValMetAspGlyValIleSe 479  
DB 1419 ACCATGAATCCAAACTCCTGAATGGTTC-CAGCGGGTGGTGTGAGCGCGTAATCTC 1477  
QY 479 rAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspGl 499  
DB 1478 TGACCAACGAGTGTGAGGAGCTGCAGAGACTGCACATGTGGCAGCAACCTCAGAGATGG 1537  
QY 499 yTyrArgGlyGlnThrSerProHisThrProAsnGluLysPheTyrGlyValThrValPh 519  
DB 1538 CTACCGGGTTCAGACCTCCCACTACTCCCAATGAAAAGTTCTATGGTGTCACTGTCTT 1597  
QY 519 eLysAlaLeuLysLeuGlyGlnGluGlyLysValProLeuGlnSerAlaHisLeuTyrTy 539  
DB 1598 CAAGCCCTCAAGCTGGGCAAGAGGCAAGTTCTCTGCAAGATGCCACCTGTACTA 1657  
QY 539 rAsnValThrGluLysValArgArgIleMetGluSerTyrPheArgLeuAspThrProLe 559  
DB 1658 CAACGTGACGGAAGGTGGCGGCATCATGTAGTCTTCCGCTTGGATACGCCCT 1717  
QY 559 uTyrPheSerTyrSerHisLeuValCysArgThrAlaIleGluGluValGlnAlaGluAr 579  
DB 1718 CTACTTTTCTTACTCTCATCTGTGTGGCCACTGCCATCGAAGAGGTCCAGGAGAGAG 1777  
QY 579 glyAspAspSerHisProValHisValAspAsnCysIleLeuAsnAlaGluThrLeuVa 599  
DB 1778 GAAGCATGATAGTATCATCCAGTCCAGTGCACAACTGCATCTGAAATGCCGAGACCTCGT 1837  
QY 599 lCysValLysGluProProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAs 619  
DB 1838 GTGTGTCAAAGAGCCCGCAGCTACACCTTCGGGACTACAGCGCCATCTTTTACCTAAA 1897  
QY 619 nGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAl 639  
DB 1898 TGGGCACTTCGATGGCGGAACCTTTTATTTCATGAATGGATGCCAGACCGTGACGGC 1957  
QY 639 aGluValGlnProGlnCysGlyArgAlaValAlaGlyPheSerSerGlyThrGluAsnProHi 659  
DB 1958 AGAGTGCAGCCTCAGTGTGGAAGAGCGGTGGGATTTCTTTCAGGCATGAAAACCCACA 2017  
QY 659 sGlyValLysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTyrPheThrLeuAs 679  
DB 2018 TGGAGTGAAGGTGTCCACAGGGGCGAGCTGTGCCATCGCCCTGTGTTCACCTCGGA 2077  
QY 679 pProArgHisSerGluArg 685  
DB 2078 CCTCGACACAGCAGCGCG 2096  
RESULT 15  
BD158280 2993 bp DNA linear PAT 17-JAN-2003  
LOCUS  
DEFINITION  
Primer for synthesizing full-length cDNA and use thereof.  
ACCESSION  
BD158280  
VERSION  
BD158280.1 GI:27864038  
KEYWORDS  
JP 2002191363-A/13123.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 2993)  
AUTHORS  
Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,  
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.  
TITLE  
Primer for synthesizing full-length cDNA and use thereof  
JOURNAL  
Patent: JP 2002191363-A 13123 09-JUL-2002;  
HELIX RESEARCH INSTITUTE  
COMMENT  
CS Homo sapiens (human)  
FN JP 2002191363-A/13123  
PD 09-JUL-2002  
PF 28-JUL-2000 JP 2000280990

PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU  
PI SAITO,  
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
PI KEIICHI NAGAI, TETSUJI OTSUKI  
PC  
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC  
10, C12P21/02, C12Q1/68, C12P21/08, C06F17/30, C12N15/00, C12N5/00 CC  
Primer for synthesizing full-length cDNA and use thereof PH Key  
Location/Qualifiers  
FT CDS  
10. .2993  
Location/Qualifiers  
1. .2993  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606".

## FEATURES

source

## ORIGIN

Alignment Scores:  
Pred. No.: 2,59e-314 Length: 2993  
Score: 295.00 Matches: 681  
Percent Similarity: 98.84% Conservative: 0  
Best Local Similarity: 98.84% Mismatches: 4  
Query Match: 40.08% Indels: 8  
DB: 6 Gaps: 0

US-10-045-815-4 (1-736) x BD158280 (1-2993)

Qy 1 MetAlaValArgAlaLeuLysLeuLeuThrThrLeuLeuAlaValAlaAlaSer 20  
Db 42 ATGGCGGTACGCGGTGGAAGCTGCTGACCACTGTGCGCTGTGCGCGCTGCGCTCC 101  
Qy 21 GlnAlaGluValGluSerGluAlaGlyTrpGlyMetValThrProAspLeuLeuPheAla 40  
Db 102 CAAGCCGAGTCGAGTCGAGCGAGGATGGGGCATGTGACGCTGATCTGCTTCGCC 161  
Qy 41 GluGlyThrAlaAlaTyAlaArgGlyAspTrpProGlyValValLeuSerMetGluArg 60  
Db 162 GAGGGGACCGCAGCTACGCGCGGGGACTGGCGCGGGTGTCTCCTGAGCATGGAACGG 221  
Qy 61 AlaLeuArgSerArgAlaAlaLeuArgAlaLeuArgLeuArgCysArgThrGlnCysAla 80  
Db 222 GCGTGTGCGCTCCCGGGAGCCCTCCGCGCCCTTCGCGCTGCGCTGCGCGCCAGCGGTGCC 281  
Qy 81 AlaAspPheProTrpGluLeuAspProAspTrpSerProAlaGlnAlaSerGly 100  
Db 282 GCGGACTTCCGCTGGAGCTGGAGCCCGACTGGTCCCGCCAGCCCGCGCGCTCGGGC 341  
Qy 101 AlaGly-AlaLeuArgAspLeuSerPhePheGlyGlyLeuLeuArgAlaAlaCysLe 120  
Db 342 GCGGC- GCGCGCTGCGCGACCTGAGCTTCTTCGGGGGCGCTTCGCGCTGCGCTGCGCT 400  
Qy 120 uArgArgCysLeuGlyProPrcAlaAlaHisSerLeuSerGluGluMetGluLeuGluPh 140  
Db 401 GCGCGCTGCTGCG 460  
Qy 140 eArgLysArgSerProTyAsnTyLeuGlnValAlaTyPheLysLeuLeuGlu 160  
Db 461 CGCGAAGCGGAGCGCCCTCAACTACCTGAGCTGCGCTTCTCAAGATCAACAAGTTGGA 520  
Qy 160 uLysAlaValAlaAlaAlaHisThrPhePheValGlyAsnProGluHisMetGluMetG1 180  
Db 521 GAAAGCTGTGTGCGAGCACACACCTTCTTGTGGGCAATCTGAGACATGGAAATGCA 580  
Qy 180 nGlnAsnLeuAspTyTrpGlnThrMetSerGlyValLysGluAlaAspPheLysAspLe 200  
Db 581 GCAGAACCTAGACTATTACCAAAACCCTGCTGGAGTGAAGGAGCGCGACTTCAAGGATCT 640  
Qy 200 uGluThrGlnProHisMetGlnGluPheArgLeuGlyValArgLeuTyTrpSerGluGlu 220  
Db 641 TGAGACTCAACCCCATATTCGAAGAATTTCGACTGGAGTGGACTCTACTCAGAGGAACA 700  
Qy 220 nProGlnGluAlaValProHisLeuGluAlaAlaLeuGlnGluTyTrpPheValAlaTyG1 240

Db 701 GCCACAGGAAGCTGTGCCCACTAGAGCGCGCTGCAAGAAATCTTTTGGCCCTATGA 760  
Qy 240 uGluCysArgAlaLeuCysGluGlyProTyAspTyTrpAspGlyTyAsnTyLeuGluTy 260  
Db 761 GGAGTGGCGCTGCGCTTCGCGAAGGGCCCTATGACTACGATGGCTCAACATCCTTGAGTA 820  
Qy 260 rAsnAlaAspLeuPheGlnAlaLeuThrAspHisTyTrpGlnValLeuAsnCysLysG1 280  
Db 821 CAACGCTGACCTTCTCCAGGCCATCACAGATCATTACATCCAGGTCCTCACTGTAAGCA 880  
Qy 280 nAsnCysValThrGluLeuAlaSerHisProSerArgGluLysProPheGluAspPheLe 300  
Db 881 GAACGTGTCACGGAGCTTGTCCCAACCAAGTCGAGAGAAGCCCTTTGAAGACTTCT 940  
Qy 300 uProSerHisTyAsnTyLeuGlnPheAlaTyTrpAsnLleGlyAsnTyTrpGlnAl 320  
Db 941 CCATCGCATTAATATCTGAGTTTGCCTACTATAACATTTGGAAATATACACAGC 1000  
Qy 320 aGly-GluCysAlaLysThrTyLeuLeuPhePheProAsnAspGluValMetAsnGlnA 340  
Db 1001 TGT-TGAATGTGCCAAGACTATCTTCTTCTTCCCAATGCGAGGTGATGAACCAA 1059  
Qy 340 snLeuAlaTyTrpAlaAlaMetLeuGlyGluGluHisThrArgSerLleGlyProArgG 360  
Db 1060 ATTGGCTTATTATGAGCTATGCTTGGAGAAGAACACACACAGATCCATCGGCCCGCTG 1119  
Qy 360 lUsrAlaLysGluTyArgGlnArgSerLeuLeuGluLysGluLeuLeuPhePheAlaT 380  
Db 1120 AGAGTCCCAAGAGATACCGACGAGCCTTACTTGGAAAAGAACTGCTTTTCTCGCT 1179  
Qy 380 yTrpValPheGlyLleProPheValAspProAspSerTrpThrProGlu-GluValLle 399  
Db 1180 ATGATGTTTTTGAATTTCTTGTGATCCGGATTCATGGATCCAGG-AGAAGTGATT 1238  
Qy 400 ProLysArgLeuGlnGlnLysGlnLysSerGluArgGluThrAlaValArgLleSerGln 419  
Db 1239 CCACAGAGATTGCAAGAGAAACAGAGTCAAGACCGGAAACACCGATCGCATCTCCAG 1298  
Qy 420 GluLleGlyAsnLeuMetLysGluLleGluThrLeuValGluLysThrLysGluSer 439  
Db 1299 GAGATTGGGAACCTTATGAGGNAATCGAGACCTTGTGGAAGAGAAGACCAAGGATCA 1358  
Qy 440 LeuAspValSerArgLeuThrArgGluGlyGlyProLeuLeuTyrcLuglyLleSerLeu 459  
Db 1359 CTGGATGTGAGCAGACTGACCGGGAAGGTGGCCCTCTCTATGAAGGCATCAGTCTC 1418  
Qy 460 ThrMetAsnSerLysLeuLeuAsnGlyTyTrp-GlnArgValValMetAspGlyValLleSe 479  
Db 1419 ACCATGAACTCCAAACTCTCTGATGGTTC-CCAGCGGGTGGTGGATGGCGGCTAATCTC 1477  
Qy 479 rAspHisGluCysGlnGluLeuGlnArgLeuThrAsnValAlaAlaThrSerGlyAspG1 499  
Db 1478 TGACCACGAGTGTGAGGAGCTGCAGAGACTGACCAATGTGGCAGCAACCTCAGGAGATGG 1537  
Qy 499 yTyArgGlyGlnThrSerProHisThrProAsnGluLysPheTyTrpGlyValThrValPh 519  
Db 1538 CTACCGGGGTGAGACTCCCCACATACTCCCAATGAAAAGTCTATGGTGTACTGCTT 1597  
Qy 519 eLysAlaLeuLysLeuGlyGlnGluLysValProLeuGlnSerAlaHisLeuTyTy 539  
Db 1598 CAAAGCCCTCAAGCTGGGCAAGAGCAAGATTCCTCTGCGAGAGTGCACCCTGTA 1657  
Qy 539 rAsnValThrGluLysValArgArgLleMetGluSerTyPheArgLeuAspThrProLe 559  
Db 1658 CAAAGTGCAGGAGAGGTGCGCGCATCATGAGTCTTACTTCCGCTCGATACGCCCT 1717  
Qy 559 uTyTrpSerTyTrpSerHisLeuValCysArgThrAlaLleGluGluValGlnAlaGluAr 579  
Db 1718 CTACTTTTCTACTCTCATCTGGTGGCGCACTGCGCATCGAAGAGGTCCAGCGCAGAG 1777  
Qy 579 gLysAspAspSerHisProValHisValAspAsnCysLleLeuAsnAlaGluThrLeuVa 599

```
Db 1778 GAAGGATGATGATCCAGTCCAGTGGACAACTGCATCCTGAATGCCGAGACCCCTCGT 1837
QY 599 ICysVallysGluProAlaTyrThrPheArgAspTyrSerAlaIleLeuTyrLeuAs 619
Db 1838 GTGTGTCAAAGAGCCCCCAGGCTACACCTCCGGGACTACAGCGCCATCCTTTACCTAAA 1897
QY 619 nGlyAspPheAspGlyGlyAsnPheTyrPheThrGluLeuAspAlaLysThrValThrAl 639
Db 1898 TGGGACTTCGATGGCGGAAACTTTATTTCACCTGAAGTGGATGCCAAGACCGTGACGGC 1957
QY 639 aGluValGlnProGlnCysGlyArgAlaValGlyPheSerSerGlyThrGluAsnProHi 659
Db 1958 AGAGGTGAGGCTCAGTGTGAAGAGCGCGTGGGATTCTCTTCAGGCACGTGAAACCCACA 2017
QY 659 sGlyVallysAlaValThrArgGlyGlnArgCysAlaIleAlaLeuTrpPheThrLeuAs 679
Db 2018 TGGAGTGAAGGCTGTACACAGGGGGCAGCGCTGTGCCATCGCCCTGTGGTTCAACCTTGA 2077
QY 679 pProArgHisSerGluArg 685
Db 2078 CCTCGACACAGCGAGCGG 2096
```

Search completed: July 18, 2004, 21:38:07  
Job time : 10818 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 17, 2004, 10:09:10 ; Search time 56 Seconds  
(without alignments)  
3713.481 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 736  
Sequence: 1 MAVRALKLLTLLAVAAAS.....PPPEAQESLSGSESKPKDEL 736

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 122046

Minimum DB seq length: 0  
Maximum DB seq length: 8

Post-processing: Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	6	0.8	7	4	AAB72810	Aab72810 Antibacte
2	6	0.8	7	4	AAB72814	Aab72814 Antibacte
3	6	0.8	7	4	AAB72822	Aab72822 Antibacte
4	6	0.8	8	2	AAR12310	Aar12310 ID2 plasm
5	6	0.8	8	2	AAR43245	Aar43245 Cytotoxic
6	6	0.8	8	2	AAR78840	Aar78840 P. falcip
7	6	0.8	8	2	AAW54563	Aaw54563 Peptide f
8	6	0.8	8	2	AAV03662	Aay03662 Amino aci
9	6	0.8	8	2	AAB23655	Aab23655 Cytotoxic
10	6	0.8	8	7	ABE80013	Abes80013 Malaria C
11	5	0.7	5	1	AAE90014	Aap90014 Peptide f
12	5	0.7	5	2	AAW00426	Aaw00426 Interleuk
13	5	0.7	5	3	AAV63463	Aay63463 Protocadn
14	5	0.7	5	3	AAE96279	Aay96279 Linker us
15	5	0.7	5	3	AAE21929	Aae21929 Oligopept
16	5	0.7	5	5	AAE21931	Aae21931 Oligopept
17	5	0.7	5	5	ABE46172	Abes46172 Desmoglei
18	5	0.7	6	1	AAE30276	Aap30276 Sequence
19	5	0.7	6	2	AAR11592	Aar11592 Fibrinoge
20	5	0.7	6	2	AAR10402	Aar10402 Fibrinoge
21	5	0.7	6	2	AAE57812	Aae57812 RGD contg
22	5	0.7	6	2	AAE57813	Aae57813 RGD contg
23	5	0.7	6	2	ADD94949	Add94949 Platelet
24	5	0.7	6	2	ADD94957	Add94957 Platelet
25	5	0.7	6	2	ADD94948	Add94948 Platelet

26	5	0.7	6	2	ADD94950	Add94950 Platelet
27	5	0.7	6	2	ADD94945	Add94945 Platelet
28	5	0.7	6	2	ADD94955	Add94955 Platelet
29	5	0.7	6	2	ADD94958	Add94958 Platelet
30	5	0.7	6	2	ADD94956	Add94956 Platelet
31	5	0.7	6	2	ADD94959	Add94959 Platelet
32	5	0.7	6	2	ADD94951	Add94951 Platelet
33	5	0.7	6	2	ADD94954	Add94954 Platelet
34	5	0.7	6	2	ADD94953	Add94953 Platelet
35	5	0.7	6	2	ADD94947	Add94947 Platelet
36	5	0.7	6	2	ADD94946	Add94946 Platelet
37	5	0.7	6	2	ADD94952	Add94952 Platelet
38	5	0.7	6	2	AAE95089	Aae95089 Silk fibr
39	5	0.7	6	2	AAW32886	Aaw32886 Peptide B
40	5	0.7	6	2	AAW26329	Aaw26329 Silk fibr
41	5	0.7	6	2	AAW64612	Aaw64612 Human poi
42	5	0.7	6	2	AAW53502	Aaw53502 Peptide r
43	5	0.7	6	2	AAW57229	Aaw57229 OA-519 pe
44	5	0.7	6	2	AAW23829	Aaw23829 Human thr
45	5	0.7	6	2	AAW96753	Aaw96753 Peptide f

ALIGNMENTS

RESULT 1  
AAB72810 ID AAB72810 standard; peptide; 7 AA.  
XX  
AC AAB72810;  
XX  
XX 09-MAY-2001 (first entry)  
XX  
DE Antibacterial peptide SEQ ID NO: 31.  
XX  
XX Antibacterial compound; Gram-negative bacterium; pilus; chaperone;  
KW biofilm; disease treatment; bacterial infection.  
XX  
OS Synthetic.  
XX  
XX WO200110386-A2.  
XX  
PD 15-FEB-2001.  
XX  
PF 11-AUG-2000; 2000WO-US022087.  
XX  
PR 11-AUG-1999; 99US-0148280P.  
XX  
XX (UNIW ) UNIV WASHINGTON.  
XX  
XX Hultgren SJ, Sauer FG, Waksman G, Fuetterer K;  
XX  
XX WPI; 2001-226496/23.  
XX  
XX An isolated compound for inhibiting pilus assembly.  
XX  
XX Claim 12; Page 138; 144pp; English.

The present invention provides antibacterial compounds which are able to interfere with Gram-negative bacteria pilus formation and assembly, and pilus interaction with chaperone proteins. These are useful in the treatment of bacterial infection, and in the prevention of biofilm formation. They are particularly useful against Escherichia coli, Haemophilus influenzae, Salmonella enteritidis, S. typhimurium, Bordetella pertussis, Yersinia enterocolitica, Helicobacter pylori and Klebsiella pneumoniae

XX SQ Sequence 7 AA;

Query Match 0.8%; Score 6; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 305 NYLQFA 310  
 DB 1 NYLQFA 6

RESULT 2  
 AAB72814  
 ID AAB72814 standard; peptide; 7 AA.  
 AC AAB72814;  
 XX  
 DT 09-MAY-2001 (first entry)  
 DE Antibacterial peptide SEQ ID NO: 35.  
 XX Antibacterial compound; Gram-negative bacterium; pilus; chaperone;  
 KW biofilm; disease treatment; bacterial infection.  
 XX Synthetic.  
 OS  
 XX WO200110386-A2.  
 XX  
 PD 15-FEB-2001.  
 XX  
 PF 11-AUG-2000; 2000WO-US022087.  
 PR 11-AUG-1999; 99US-0148280P.  
 XX  
 PA (UNIW ) UNIV WASHINGTON.  
 XX  
 PI Hultgren SJ, Sauer FG, Waksman G, Fuetterer K;  
 XX WPI; 2001-226496/23.  
 DR  
 XX An isolated compound for inhibiting pilus assembly.  
 PT  
 XX Claim 12; Page 141; 144pp; English.  
 PS  
 XX The present invention provides antibacterial compounds which are able to  
 CC interfere with Gram-negative bacteria pilus formation and assembly, and  
 CC pilus interaction with chaperone proteins. These are useful in the  
 CC treatment of bacterial infection, and in the prevention of biofilm  
 CC formation. They are particularly useful against *Escherichia coli*,  
 CC *Haemophilus influenzae*, *Salmonella enteritidis*, *S. typhimurium*,  
 CC *Bordetella pertussis*, *Yersinia enterocolitica*, *Helicobacter pylori* and  
 CC *Klebsiella pneumoniae*  
 CC  
 XX Sequence 7 AA;  
 SQ

Query Match 0.8%; Score 6; DB 4; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 305 NYLQFA 310  
 DB 1 NYLQFA 6

RESULT 4  
 AAR12310  
 ID AAR12310 standard; protein; 8 AA.  
 XX  
 AC AAR12310;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 29-AUG-1991 (first entry)  
 XX  
 DE ID2 plasmodium surface peptide.  
 XX  
 KW Immunogenic determinant; circumsporozoite; CS; vaccine; malaria; hybrid.  
 XX  
 OS Plasmodium falciparum.  
 XX  
 PN EP432965-A.  
 XX  
 PD 19-JUN-1991.  
 XX  
 PF 06-DEC-1990; 90EP-00313257.  
 XX  
 PR 08-DEC-1989; 89US-00447746.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM.  
 PA (USSA ) US SEC OF ARMY.  
 PA (BIOM-) BIOMEDICAL RES INST.  
 PA (GROS/) GROSS M S.  
 XX  
 PI Gross MS, Gordon DM, Hollingdal MR;  
 XX WPI; 1991-179771/25.  
 DR  
 XX Polypeptide comprising immunogenic determinants from *P. falciparum* - for  
 PT vaccine against malaria infection in humans.  
 XX

QY 305 NYLQFA 310  
 DB 1 NYLQFA 6

RESULT 3  
 AAB72822  
 ID AAB72822 standard; peptide; 7 AA.  
 AC AAB72822;  
 XX  
 DT 09-MAY-2001 (first entry)  
 DE Antibacterial peptide SEQ ID NO: 43.  
 XX  
 KW Antibacterial compound; Gram-negative bacterium; pilus; chaperone;  
 KW biofilm; disease treatment; bacterial infection.  
 XX  
 OS Synthetic.  
 XX

Query Match 0.8%; Score 6; DB 4; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 305 NYLQFA 310  
 DB 1 NYLQFA 6

RESULT 3  
 AAB72822  
 ID AAB72822 standard; peptide; 7 AA.  
 AC AAB72822;  
 XX  
 DT 09-MAY-2001 (first entry)  
 DE Antibacterial peptide SEQ ID NO: 43.  
 XX  
 KW Antibacterial compound; Gram-negative bacterium; pilus; chaperone;  
 KW biofilm; disease treatment; bacterial infection.  
 XX  
 OS Synthetic.  
 XX



PS Claim 2; Page 16; 18pp; English.

XX The peptide is a sporozoite neutralising epitope from the 2nd flanking region of the CS protein of plasmodium. It can be used in a vaccine for protection against malaria. The peptide is pref. linked to a 2nd peptide from the 1st flanking domain and the resulting polypeptide fused to a carrier protein, e.g. tetanus toxoid, dip-theria toxin or cholera B toxin. A preferred vaccine comprises 81 N-terminal AAs of the influenza virus nonstructural protein 1 (NS181), fused, via a synthetic linker, to a 1st flanking region of the CS protein, which is itself fused to a 2nd flanking region of the CS protein. See also AAR12306-R12311 and AAR13175-R13179. (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 8 AA;

SQ Query Match 0.8%; Score 6; DB 2; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 731 KPDEL 736  
| | | | |  
Db 2 KPDEL 7

RESULT 5

AAR43245

ID AAR43245 standard; peptide; 8 AA.

AC AAR43245;

XX 25-MAR-2003 (revised)

DT 04-MAY-1994 (first entry)

XX Cytotoxic T lymphocyte recognition/induction peptide.

DE CTL; vaccine; malaria; specific antigen-derived.

XX Synthetic.

OS WO9320103-A2.

XX 14-OCT-1993.

PD 05-APR-1993; 93WO-GB000711.

XX 03-APR-1992; 92GB-00008068.

PR 20-AUG-1992; 92GB-00017704.

XX (ISIS-) ISIS INNOVATION LTD.

PA Hill AV, Gotch FM, McMichael AJ, Whittle HC;

XX WPI; 1993-336833/42.

DR Peptide(s) recognising or inducing cytotoxic T lymphocytes - useful in vaccines against malaria or HIV-2, derived from specific antigen and human leukocyte antigen contg. class I restricted epitope.

XX Claim 6; Page 30; 35pp; English.

PS The sequence is that of peptide cp26 which is recognised by, or can induce, cytotoxic T lymphocytes. It may be useful in vaccines against malaria. (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 8 AA;

SQ Query Match 0.8%; Score 6; DB 2; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 731 KPDEL 736  
| | | | |  
Db 1 KPDEL 6

RESULT 6

AAR78840

ID AAR78840 standard; peptide; 8 AA.

XX AAR78840;

AC AAR78840;

XX 25-MAR-2003 (revised)

DT 27-MAR-1996 (first entry)

XX P. falciparum circumsp. 368-375 cytotoxic T lymphocyte epitope.

XX Circump. 368-375; cytotoxic T; CTL; epitope; helper T; HTL; cell; lymphocyte; viruses; parasites; tumours; antigens; treatment; disease prevention.

XX Plasmodium falciparum.

OS WO9522317-A1.

PN 24-AUG-1995.

XX 16-FEB-1995; 95WO-US002121.

PF 16-FEB-1994; 94US-00197484.

XX (CYTE-) CYTEL CORP.

PA Vitello MA, Chesnut RW, Sette AD, Celis E, Grey H;

PI WPI; 1995-302545/39.

DR Compn. inducing cytotoxic T lymphocyte response to pref. viral, bacterial, parasitic or tumour antigens - useful in the treatment and prevention of diseases associated with the antigen e.g. hepatitis B.

XX Disclosure; Page 17; 109pp; English.

PS A compn. which induces a cytotoxic T lymphocyte (CTL) response to an antigen (Ag) in a mammal comprises, a CTL Ag response inducing peptide (i.e. AAR78824-R78853) and a lipid conjugated helper T cell inducing peptide. The compn. induces a CTL response to bacterial, viral or tumour Ags, and is therefore useful in the treatment and prevention of diseases associated with the Ag. (Updated on 25-MAR-2003 to correct PI field.)

XX Sequence 8 AA;

SQ Query Match 0.8%; Score 6; DB 2; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 731 KPDEL 736  
| | | | |  
Db 1 KPDEL 6

RESULT 7

AAW54663

ID AAW54663 standard; peptide; 8 AA.

XX AAW54663;

AC AAW54663;

XX 25-SEP-1998 (first entry)

DT Peptide from P.falciparum CSP 368-375.

XX Mannose; antigen; antigen-presenting cell; mannosylated peptide; T cell; vaccine; treatment.

XX Synthetic.

OS WO9813378-A1.

PN

XX 02-APR-1998.  
 PD  
 XX  
 XX 25-SEP-1997; 97WO-NL000536.  
 PF  
 XX  
 XX 26-SEP-1996; 96EP-00202701.  
 PR  
 XX  
 XX (UYLE-) RIJKSUNIV LEIDEN.  
 PA  
 XX  
 PI Koning F, Drijfhout JW;  
 PI  
 XX WPI; 1998-230631/20.  
 DR  
 XX  
 XX Increasing uptake and presentation of antigen(s) - by adding mannose  
 PT residue(s) to antigen for increasing T cell response, useful in, e.g.  
 PT vaccines against viral infection(s).  
 PT  
 XX  
 XX Disclosure; Page 29; 47pp; English.  
 PS  
 XX The peptides AAW54559-W54809 are examples of peptides to which at least 1  
 CC (preferably 2) mannose can be attached to increase their uptake as  
 CC antigens by antigen-presenting cells. Uptake of agonist mannoseylated  
 CC peptides will increase the T cell response, whereas uptake of antagonist  
 CC peptides blocks the T cell response. Blocking binding of immunogenic  
 CC autoantigens can be used in treatment of type I diabetes, rheumatoid  
 CC arthritis, graft rejection etc., also to induce T-cell non-  
 CC responsiveness. Vaccines containing mannoseylated antigen are used to  
 CC prevent or treat infections by, e.g. bacteria, viruses, fungi, helminths  
 CC and parasites  
 CC  
 XX Sequence 8 AA;  
 SQ  
 Query Match 0.8%; Score 6; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 731 KPKDEL 736  
 Db |||||  
 1 KPKDEL 6  
 RESULT 8  
 AAY03662  
 ID AAY03662 standard; peptide; 8 AA.  
 XX  
 XX AAY03662;  
 AC  
 XX  
 DT 07-JUN-1999 (first entry)  
 DE  
 DE Amino acid sequence of the malaria (M) string CTL epitope Cp26.  
 XX  
 XX CD8+ T-cell; immune response; antigen; priming composition; CTL; epitope;  
 KW cytotoxic T lymphocyte; boosting; poxvirus vector; PVV; pathogen; tumour;  
 KW malaria; parasite; P. falciparum; viral; bacterial; parasitic; cancer;  
 KW melanoma; HIV; breast; colon; vaccination.  
 XX  
 OS Plasmodium falciparum.  
 XX  
 XX WO9856919-A2.  
 PN  
 XX  
 XX 17-DEC-1998.  
 PD  
 XX  
 XX 09-JUN-1998; 98WO-GB001681.  
 PF  
 XX  
 XX 09-JUN-1997; 97GB-00011957.  
 PR  
 XX  
 XX (ISIS-) ISIS INNOVATION LTD.  
 PA  
 XX  
 XX McMichael AJ, Hill AVS, Gilbert SC, Schneider J, Plebanski M;  
 PI Hanke T, Smith GL, Blanchard T;  
 PI  
 XX  
 XX WPI; 1999-070325/06.  
 DR  
 DR N-PSDB; AAX29202.

XX  
 PT Generating CD8-positive T cell response to target antigen using  
 PT recombinant poxvirus - for treating or preventing malaria and HIV  
 PT infection, also epitope strings from Plasmodium and HIV.  
 XX  
 XX Claim 38; Page 18; 85pp; English.  
 PS  
 XX The invention relates to methods and reagents for generating a protective  
 CC CD8+ T-cell immune response against at least one target antigen. The kits  
 CC of the invention comprises (i) as priming composition, a source of one or  
 CC more CD8+ T-cell [cytotoxic T lymphocytes-(CTL)] epitopes of the target  
 CC antigen, plus a carrier and (ii) as boosting composition a source of CTL  
 CC epitopes, with at least one CTL epitope the same as used in (i), with  
 CC this source being a non-replicating or replication-impaired recombinant  
 CC poxvirus vector (PVV) plus a carrier. If the source of CTL epitopes in  
 CC (i) is a viral vector, then the vector in (ii) is from a different virus.  
 CC The kits are used to generate an immune response (prophylactic or  
 CC therapeutic) against pathogens or tumours, specifically against malaria  
 CC parasites such as P. falciparum, or HIV, and also many other bacterial,  
 CC viral or parasitic pathogens. The kits are also used for protective  
 CC response against melanoma and cancer of breast or colon, and generally  
 CC wherever a strong CD8+ response is protective. The boosting composition  
 CC may be used alone to boost a naturally primed response against malaria.  
 CC The specified PVV provide an excellent booster effect, better than that  
 CC from wild-type poxvirus, resulting in complete rather than partial  
 CC protection against sporozoite challenge. Also PVV are safer to use than  
 CC wild-type virus. Sequences AAY03661-680 represent CTL peptide epitopes of  
 CC the malaria (M) string  
 XX  
 SQ Sequence 8 AA;  
 Query Match 0.8%; Score 6; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 731 KPKDEL 736  
 Db |||||  
 1 KPKDEL 6  
 RESULT 9  
 AAB23655  
 ID AAB23655 standard; peptide; 8 AA.  
 XX  
 XX AAB23655;  
 AC  
 XX  
 DT 05-JAN-2001 (first entry)  
 DE  
 DE Cytotoxic T lymphocyte (CTL) epitope SEQ ID NO:7.  
 XX  
 XX ATPase; Hsp70; heat shock protein; cytotoxic T lymphocyte; CTL;  
 KW immune response; infectious disease; malaria; cytotoxic T cell;  
 KW cytostatic; immunostimulant; cellular immune response inducer;  
 KW protozoacide; leukaemia; cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200049041-A1.  
 PN  
 XX  
 XX 24-AUG-2000.  
 PD  
 XX  
 XX 18-FEB-2000; 2000WO-JP000941.  
 PF  
 XX  
 XX 19-FEB-1999; 99JP-00041535.  
 PR  
 XX  
 XX (SUME) SUMITOMO ELECTRIC IND CO.  
 PA  
 XX  
 XX Shinbara N, Udono H, Yui K;  
 PI  
 XX WPI; 2000-543748/49.  
 DR  
 XX  
 XX Fused protein capable of inducing cellular immune response, useful as  
 PT active ingredient for drug compositions in preventing and/or treating

PT infectious diseases such as malaria or cancer.

PS Claim 5; Page 52; 72pp; Japanese.

XX The present invention describes a fused protein (I) prepared from a peptide containing a CTL (cytotoxic T lymphocyte) epitope recognised by cytotoxic T cells and a protein containing the ATPase domain of a heat shock protein. Also described are: (1) a drug composition containing (I) as active ingredient; (2) a DNA encoding (I); (3) an expression vector containing the DNA of (2); and (4) a transformant which can retain the expression vector of (3). (1) has cytostatic, immunostimulant and protozoacide activities, and can be used as a cellular immune response inducer. The protein is useful as an active ingredient for drug compositions in preventing and/or treating infectious diseases such as malaria or cancer e.g. to provide systemic immunity against leukaemia. The present sequence represents a specifically claimed CTL epitope for use in a fused protein of the present invention

XX SQ Sequence 8 AA;

Query Match 0.8%; Score 6; DB 3; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPKDEL 736  
Db 1 KPKDEL 6

RESULT 10

AD890013  
ID ADE80013 standard; peptide; 8 AA.

AC ADE80013;

DT 29-JAN-2004 (first entry)

XX Malaria CTL epitope in method to generate CD8+ T-cell immune response.

XX antimalarial; cytostatic; vaccine; immune response;

KW non-hepadnaviral antigen; hepatitis B core particle; CD8+ T-cell;

KW epitope; poxvirus vector; cancer; malaria; epitope.

XX Plasmodium falciparum.

XX WO2003066833-A2.

XX 14-AUG-2003.

XX 07-FEB-2003; 2003WO-US003897.

XX 08-FEB-2002; 2002US-0354963P.

XX (UUNY-) UNIV NEW YORK MEDICAL CENT.

XX Zavala F, Birkett AJ;

XX WPI; 2003-748124/70.

XX Generating an immune response against a non-hepadnaviral antigen in a mammal, useful for treating or preventing cancer or malaria, by administering a priming component comprising a recombinant hepatitis B core particle.

PS Disclosure; SEQ ID NO 69; 85pp; English.

XX The invention relates to a method of generating an immune response against a non-hepadnaviral antigen in a mammal by administering (to the mammal) at least 1 dose of a priming component comprising a recombinant hepatitis B core particle (RHEP) (which is a carrier for 1 or more non-hepadnaviral CD8+ T-cell epitopes of the antigen). The method may be supplemented by the use of a boosting stage comprising a non-replicating or replication-impaired recombinant poxvirus vector. The method is useful

CC for generating an immune response against a non-hepadnaviral antigen in a mammal for treating or preventing cancer or malaria. This sequence represents a Plasmodium falciparum CTL peptide used to generate an immune response against a Plasmodium peptide.

XX SQ Sequence 8 AA;

Query Match 0.8%; Score 6; DB 7; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPKDEL 736  
Db 1 KPKDEL 6

RESULT 11

AD900014  
ID AAP90014 standard; protein; 5 AA.

XX AAP90014;

XX 25-MAR-2003 (revised)

DT 01-NOV-1989 (first entry)

XX Peptide from HLA Class II beta region contg. residue corresp. tp DQ-beta protein.

XX Peptide; DQ-beta locus of HLA class II beta genes; allele-specific.

OS Homo sapiens.

XX WO8904875-A.

XX 01-JUN-1989.

XX 14-NOV-1989; 89WO-US004067.

XX 17-NOV-1987; 87US-00121519.

XX (CETU) CETUS CORP.

XX Erlich HA, Horn GT;

XX WPI; 1989-178393/24.

XX Marker DNA sequences from HLA class-II beta region - detect amino acid 57 codon of dq-beta protein to detect auto-immune susceptibility.

XX Claim 14; Page 59; 72pp; English.

XX Peptide contg. an epitope which has an amino acid residue corresp. to position 57 of a DQ-beta protein from the HLA class II beta genes, used to raise antibodies to (indirectly) detect the identity of codon-57 of the DQ-beta protein sequence. Pref. codon-57 is selected from Ala, Val and Asp. Used to detect autoimmune diseases, esp. diabetes mellitus, and Pemphigus vulgaris. (Updated on 25-MAR-2003 to correct PF field.)

XX SQ Sequence 5 AA;

Query Match 0.7%; Score 5; DB 1; Length 5;

Best Local Similarity 100.0%; Pred. No. 1.4e+06;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 GPPAA 129

Db 1 GPPAA 5

RESULT 12

AAW00426

ID AAW00426 standard; peptide; 5 AA.

XX

AC AAW00426;  
 XX 29-AUG-1996 (first entry)  
 DT XX  
 DE Interleukin-6 antagonist peptide.  
 XX IL-6; antagonist; autoimmune disease.  
 XX Synthetic.  
 OS  
 XX JF07324097-A.  
 PN 12-DEC-1995.  
 XX  
 PD 30-MAY-1994; 94JP-00117259.  
 XX  
 PF 30-MAY-1994; 94JP-00117259.  
 XX  
 PR (DAIL ) DAICEL CHEM IND LTD.  
 XX (FUJI ) FUJISAWA PHARM CO LTD.  
 PA  
 PP WPI; 1996-065476/07.  
 XX  
 PS Interleukin 6 antagonist - useful for treating auto-immune diseases.  
 PT  
 XX Claim 3, 6; Page 2, 3; 19pp; Japanese.  
 PS  
 CC New IL-6 antagonists are provided which are of formula X-W-Y, in which X  
 CC is H or an amino-protecting group; Y is OH or a carboxy- protecting  
 CC group, and W is a peptide containing all or part of the sequence as given  
 CC in AAW00401, AAW00402, AAW00403 or AAW00404, where any free mercapto  
 CC groups in the sequence are optionally protected. The present sequence is  
 CC a specifically preferred partial sequence of AAW00402 and is itself  
 CC claimed as a new chemical entity. The IL-6 antagonists are useful for  
 CC treating autoimmune diseases  
 XX  
 SQ Sequence 5 AA;  
 Query Match 0.7%; Score 5; DB 2; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 68 LRALR 72  
 Db |||||  
 1 LRALR 5  
 RESULT 13  
 AAV63463  
 ID AAV63463 standard; peptide; 5 AA.  
 XX  
 AC AAV63463;  
 XX  
 DT 02-MAR-2000 (first entry)  
 XX  
 DE Protocadherin cell adhesion recognition cyclic peptide SEQ ID NO:2953.  
 XX  
 KW Modulation; nonclassical cadherin mediated cell adhesion; CAR;  
 KW inhibition; cadherin extracellular domain; cell adhesion recognition;  
 KW OB-cadherin; cadherin-5; cadherin-6; cadherin-7; cadherin-8; cadherin-12;  
 KW cadherin-14; cadherin-15; T-cadherin; PB-cadherin;  
 KW cadherin related neuronal receptor; LI-cadherin; protocadherin;  
 KW desmoglein; desmocollin; calcium binding; cancer; tumour; obesity;  
 KW rheumatoid arthritis; multiple sclerosis; diabetes; metastasis;  
 KW neurological disease; cyclic.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 PH Modified-site 1..5  
 FT /note= "the terminal residues are condensed with each  
 FT other to form a cyclic peptide"

XX WO9957149-A2.  
 XX 11-NOV-1999.  
 XX  
 XX 05-MAY-1999; 99WO-CA000363.  
 XX  
 XX 05-MAY-1998; 98US-00073040.  
 XX 06-NOV-1998; 98US-00187859.  
 XX 20-JAN-1999; 99US-00234395.  
 XX 08-MAR-1999; 99US-00264516.  
 XX  
 PA (ADHE-) ADHEREX TECHNOLOGIES INC.  
 XX  
 XX Blaschuk OW, Gour BJ, Byers S;  
 XX WPI; 2000-038791/03.  
 DR  
 XX New cadherin modulating agents, used for modulating nonclassical cadherin  
 PT -mediated functions for treating e.g. cancers, obesity, rheumatoid  
 PT arthritis, multiple sclerosis, diabetes or a neurological disease.  
 XX  
 PS Claim 84; Page 204; 252pp; English.  
 XX  
 CC The present invention describes cadherin modulating agents (MA)  
 CC comprising peptides which comprise a nonclassical cadherin cell adhesion  
 CC recognition (CAR) sequence. The MAs can be used for modulating  
 CC nonclassical cadherin-mediated functions. They can be used for e.g.  
 CC inhibiting adhesion of nonclassical-cadherin expressing cells in a  
 CC mammal, enhancing delivery of a drug through the skin of a mammal,  
 CC enhancing delivery of a drug to a tumour in a mammal, treating cancer in  
 CC a mammal, inhibiting metastasis of a cancer in a mammal, inhibiting  
 CC angiogenesis in a mammal, inducing apoptosis in a nonclassical cadherin-  
 CC expressing cell, preventing or treating obesity in a mammal, stimulating  
 CC blood vessel regression in a mammal, enhancing drug delivery to the  
 CC central nervous system, treating a demyelinating neurological disease,  
 CC increasing vasopermeability in a mammal, enhancing adhesion of  
 CC nonclassical cadherin-expressing cells, inhibiting synaptic stability in  
 CC a mammal, or preventing pregnancy in a mammal. They can also be used for  
 CC e.g. enhancing or directing neurite outgrowth, facilitating wound healing  
 CC or reducing scar tissue, or enhancing adhesion of foreign tissue in a  
 CC mammal. They can also be used for treating e.g. psoriasis, arthritis, age  
 CC related macular degeneration, multiple sclerosis and diabetes. The  
 CC products can also be used for detection and diagnosis and in bioreactors.  
 CC AAY60592 to AAY64572 represent specifically claimed peptides, and  
 CC AAY64573 to AAY64643 and AAZ33183 to AAZ33186 represent sequences used in  
 CC the exemplification of the present invention  
 XX  
 SQ Sequence 5 AA;  
 Query Match 0.7%; Score 5; DB 3; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 691 DDLVK 695  
 Db |||||  
 1 DDLVK 5  
 RESULT 14  
 AAY96279  
 ID AAY96279 standard; peptide; 5 AA.  
 XX  
 AC AAY96279;  
 XX  
 DT 11-AUG-2000 (first entry)  
 XX  
 DE Linker used to find IGF-1 receptor interacting proteins IIP cDNAs.  
 XX  
 DE IGF-1; IGF-1 receptor interacting protein; IIP; cancer; gene therapy;  
 KW tumour proliferation; signal transduction; protein-protein interaction;  
 KW diabetes; neurodegenerative disease; bone disease; agonist;  
 KW prostate cancer; breast cancer; linker.

```

XX OS Mammalia.
XX PN EPI006184-A1.
XX PD 07-JUN-2000.
XX PF 03-DEC-1998; 98EP-00122992.
XX PR 03-DEC-1998; 98EP-00122992.
XX PA (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX DR WPI; 2000-367969/32.
XX DR N-PSDB; AAA27351.
XX PT New IGF-1R-interacting protein for diagnosis of tumor progression and
XX treatment of tumors.
XX PS Disclosure; Fig 1; 31pp; English.
XX CC The present sequence is a peptide linker inserted between the LexA-
XX binding domain and IGF-1 receptor domain in a two-hybrid bait constructed
XX to screen a cDNA library for the mammalian IGF-1 receptor interacting
XX protein (IIP) genes. The proteins interact with IGF-1 receptor in the
XX cell, which is in turn overexpressed in many tumours. The use of the
XX proteins and genes in gene therapy and to identify agonists, antagonists
XX and antibodies will enable the diagnosis and treatment of disorders such
XX as diabetes, neurodegenerative disorders, bone diseases and cancers, in
XX particular prostate and breast cancers, as well as enabling the
XX inhibition of tumour progression. The genes are also useful for creating
XX animal models for these diseases
XX SQ Sequence 5 AA;
XX Query Match 0.7%; Score 5; DB 3; Length 5;
XX Best Local Similarity 100.0%; Pred. No. 1.4e+06;
XX Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 139 EPRKR 143
XX |||||
XX Db 1 EPRKR 5
XX
XX RESULT 15
XX ID AAE21929
XX AC AAE21929 standard; peptide; 5 AA.
XX AC AAE21929;
XX DT 16-JUL-2002 (first entry)
XX DE Oligopeptide #5 used to generate plasmin-cleavable prodrug.
XX KW plasmin; prodrug; cancer; neoplastic disease; inflammatory disease;
XX tumour; infectious disease; chemosensitiser; dye; cytostatic.
XX OS Unidentified.
XX FH Key Location/Qualifiers
XX FT Misc-difference 1 /note= "D-form residue"
XX FT FT Misc-difference 5 /note= "Optionally linked to daunorubicin (DNR) or
XX FT doxorubicin (DOX) drug"
XX PN WO200215700-A1.
XX PD 28-FEB-2002.
XX PF 23-AUG-2001; 2001WO-US026476.
XX PR 24-AUG-2000; 2000US-0227686P.

```

```

XX (COUS ) COULTER PHARM INC.
XX PI Trouet A, Dubois V, Passioukov A;
XX WPI; 2002-339507/37.
XX PT New plasmin-cleavable prodrugs comprise an oligopeptide linked to a
XX stabilizing group and a therapeutic agent.
XX PS Example 1; Page 43; 67pp; English.
XX CC The present invention relates to plasmin-cleavable prodrugs comprising an
XX oligopeptide directly linked at a first attachment site to a stabilising
XX group that hinders cleavage of the oligopeptide by enzymes present in
XX whole blood and linked at a second attachment site, either directly or
XX through a linker, to a therapeutic agent capable of entering a target
XX cell. The prodrugs of the invention are useful for delivering therapeutic
XX agents, preferably alkylating agents, anthracyclines, antiproliferative
XX agents, camptothecins, chemotherapeutic agents, cyclosporins, enediyne,
XX dolastatins, epipodophyllotoxins, maytansinoids, naphthalimides, platinum
XX complexes, pteridines, rhodamines, sulfoximides, topoisomerase inhibitors,
XX taxanes, taxoids, tubulin binding agents or vinca alkaloids, especially
XX daunorubicin or doxorubicin. They are used to target cells in high-
XX plasmin environments so that the therapeutic agent is cleaved from the
XX prodrug in a form capable of entering the target cells, especially in the
XX treatment of cancer, neoplastic diseases, inflammatory diseases, tumours
XX and infectious diseases. Sequences of the invention can be used with
XX agents that act as chemosensitisers and dyes. The present sequence is an
XX oligopeptide used to generate plasmin-cleavable prodrugs of the invention
XX SQ Sequence 5 AA;
XX Query Match 0.7%; Score 5; DB 5; Length 5;
XX Best Local Similarity 100.0%; Pred. No. 1.4e+06;
XX Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 5 ALKLL 9
XX |||||
XX Db 1 ALKLL 5
XX
XX Search completed: August 17, 2004, 10:15:20
XX Job time : 58 secs

```

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 17, 2004, 10:14:15 : Search time 19 Seconds  
(without alignments)  
1999.827 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 736  
Sequence: 1 MAVRALKLTLLAVAAAS.....PPEPAQESUSGSSSKPKDEL 736

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 71086

Minimum DB seq length: 0  
Maximum DB seq length: 8

Post-processing: Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pap.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pap.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pap.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pap.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS COMB.pap.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	0.8	8	2	US-08-318-856A-3
2	6	0.8	8	4	US-08-197-484-17
3	6	0.8	8	4	US-09-454-204A-4
4	6	0.8	8	5	PCT-US95-02121-17
5	5	0.7	5	4	US-09-187-859-2953
6	5	0.7	5	4	US-09-839-542B-2953
7	5	0.7	5	4	US-09-535-852-916
8	5	0.7	6	1	US-07-609-716-7
9	5	0.7	6	1	US-08-175-155-1
10	5	0.7	6	1	US-08-469-005A-5
11	5	0.7	6	1	US-08-211-747-4
12	5	0.7	6	1	US-08-188-426-5
13	5	0.7	6	1	US-08-477-509B-21
14	5	0.7	6	2	US-08-707-237A-6
15	5	0.7	6	2	US-08-659-486A-4
16	5	0.7	6	2	US-08-469-009-5
17	5	0.7	6	2	US-08-469-007-5
18	5	0.7	6	3	US-09-012-126-6
19	5	0.7	6	3	US-08-482-085B-21
20	5	0.7	6	3	US-08-782-480-41
21	5	0.7	6	3	US-08-336-553A-67
22	5	0.7	6	3	US-09-083-351-22
23	5	0.7	6	3	US-08-475-411A-7
24	5	0.7	6	3	US-08-478-029A-7
25	5	0.7	6	3	US-09-083-352-22
26	5	0.7	6	3	US-08-954-211-41
27	5	0.7	6	3	US-09-395-344-6

Sequence 37, Appl  
Sequence 21, Appl  
Sequence 1155, Ap  
Sequence 2, Appl  
Sequence 41, Appl  
Sequence 67, Appl  
Sequence 27, Appl  
Sequence 9, Appl  
Sequence 1155, Ap  
Sequence 41, Appl  
Sequence 2, Appl  
Sequence 923, App  
Sequence 951, App  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 55, Appl  
Sequence 2, Appl

28 5 0.7 6 4 US-09-455-679-37  
29 5 0.7 6 4 US-09-444-791A-21  
30 5 0.7 6 4 US-09-187-859-1155  
31 5 0.7 6 4 US-08-806-029-2  
32 5 0.7 6 4 US-09-005-167A-41  
33 5 0.7 6 4 US-08-439-157-67  
34 5 0.7 6 4 US-09-437-895-67  
35 5 0.7 6 4 US-09-155-613A-27  
36 5 0.7 6 4 US-09-138-277C-9  
37 5 0.7 6 4 US-09-176-741B-41  
38 5 0.7 6 4 US-09-839-542B-1155  
39 5 0.7 6 4 US-09-863-859-2  
40 5 0.7 6 4 US-09-535-852-923  
41 5 0.7 6 4 US-09-535-852-951  
42 5 0.7 7 1 US-08-469-005A-2  
43 5 0.7 7 1 US-08-188-426-2  
44 5 0.7 7 2 US-08-529-190B-55  
45 5 0.7 7 2 US-08-469-009-2

#### ALIGNMENTS

RESULT 1  
US-08-318-856A-3  
: Sequence 3, Application US/08318856A  
: Patent No. 5972351  
: GENERAL INFORMATION:  
: APPLICANT: Adrian V.S. Hill, et al.  
: TITLE OF INVENTION: PLASMODIUM FALCIPARUM MHC CLASS I-  
: TITLE OF INVENTION: RESTRICTED CTL EPITOPES DERIVED FROM PRE-ERYTHROCYTIC STAGE  
: NUMBER OF SEQUENCES: 86  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
: STREET: 2033 K Street, N.W., Suite 800  
: CITY: Washington  
: STATE: D.C.  
: COUNTRY: U.S.A.  
: ZIP: 20006  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk, 3.5 inch, 1.44 mb  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: WordPerfect 5.1+  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/318,856A  
: FILING DATE: October 3, 1994  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: GB 92 08 068.8  
: FILING DATE: April 3, 1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: GB 92 17 704.7  
: FILING DATE: August 20, 1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: WO PCT/GB93/00711  
: FILING DATE: April 5, 1993  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Lee Cheng  
: REGISTRATION NUMBER: 40,949  
: REFERENCE/DOCKET NUMBER: 263-PP1R1577US  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (202) 721-8200  
: TELEFAX: (202) 721-8250  
: INFORMATION FOR SEQ ID NO: 3:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 8 amino acid residues  
: TYPE: amino acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: peptide  
: ORIGINAL SOURCE:  
: ORGANISM: Plasmodium falciparum

US-08-318-856A-3

Query Match 0.8%; Score 6; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred.No. 3e+05;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPDEL 736

Db 1 KPDEL 6  
 |||||

RESULT 2

US-08-197-484-17  
 ; Sequence 17, Application US/08197484  
 ; Patent No. 6419931  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VITIELLO, Maria A.  
 ; APPLICANT: CHESTNUT, Robert W.  
 ; APPLICANT: SETTE, Alessandro D.  
 ; APPLICANT: CELIS, Esteban  
 ; APPLICANT: GRAY, Howard  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ELICITING  
 ; TITLE OF INVENTION: CTL IMMUNITY  
 ; NUMBER OF SEQUENCES: 153  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Townsend and Townsend Khourie and Crew  
 ; STREET: Steuart Street Tower, One Market Plaza  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: US  
 ; ZIP: 94105-1493

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/197,484  
 FILING DATE: 16-FEB-1994  
 CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/935,811  
 FILING DATE: 26-AUG-1992

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/874,491  
 FILING DATE: 27-APR-1992

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/827,682  
 FILING DATE: 29-JAN-1992

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/749,568  
 FILING DATE: 28-AUG-1991

ATTORNEY/AGENT INFORMATION:  
 NAME: Parmelee, Steven W.  
 REGISTRATION NUMBER: 31,990

REFERENCE/DOCKET NUMBER: 14137-26-4

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (206) 467-9600  
 TELEFAX: (206) 623-6793

INFORMATION FOR SEQ ID NO: 17:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 8 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: unknown  
 TOPOLOGY: unknown  
 MOLECULE TYPE: peptide

US-08-197-484-17

Query Match 0.8%; Score 6; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred.No. 3e+05;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

731 KPDEL 736

Db 1 KPDEL 6

|||||

RESULT 3

US-09-454-204A-4  
 ; Sequence 4, Application US/09454204A  
 ; Patent No. 6663871  
 ; GENERAL INFORMATION:  
 ; APPLICANT: McMichael, Andrew  
 ; APPLICANT: Hill, Adrian V.S.  
 ; APPLICANT: Gilbert, Sarah C.  
 ; APPLICANT: Schneider, Jorg  
 ; APPLICANT: Plebanski, Magdalena  
 ; APPLICANT: Hanke, Tomas  
 ; APPLICANT: Smith, Geoffrey L.  
 ; APPLICANT: Blanchard, Tom

; TITLE OF INVENTION: Methods and Reagents for Vaccination  
 ; TITLE OF INVENTION: Which Generate A CD8 T Cell Immune Response

; FILE REFERENCE: 2907.1000-000  
 ; CURRENT APPLICATION NUMBER: US/09/454,204A

; CURRENT FILING DATE: 1999-12-09  
 ; PRIOR APPLICATION NUMBER: PCT/GB98/01681

; PRIOR FILING DATE: 1998-06-09  
 ; PRIOR APPLICATION NUMBER: GB 97 11957.2

; PRIOR FILING DATE: 1997-06-09  
 ; NUMBER OF SEQ ID NOS: 78

; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4

; LENGTH: 8  
 ; TYPE: PRT

; ORGANISM: Unknown  
 ; FEATURE:  
 ; OTHER INFORMATION: CTL Epitope of the Malaria String

US-09-454-204A-4

Query Match 0.8%; Score 6; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred.No. 3e+05;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPDEL 736

Db 1 KPDEL 6  
 |||||

RESULT 4

PCT-US95-02121-17  
 ; Sequence 17, Application PC/TUS9502121  
 ; GENERAL INFORMATION:  
 ; APPLICANT:  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ELICITING  
 ; TITLE OF INVENTION: CTL IMMUNITY  
 ; NUMBER OF SEQUENCES: 153  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US95/02121  
 FILING DATE: 16-FEB-1995  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/197,484

; FILING DATE: 16-FEB-1994  
 ; PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/935,811

; FILING DATE: 26-AUG-1992  
 ; PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/874,491

; FILING DATE: 27-APR-1992  
 ; PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/827,682

; FILING DATE: 29-JAN-1992  
 ; PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/749,568

; FILING DATE: 28-AUG-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 NAME: Parmelee, Steven W.  
 REGISTRATION NUMBER: 31,990

REFERENCE/DOCKET NUMBER: 14137-26-4

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (206) 467-9600  
 TELEFAX: (206) 623-6793

INFORMATION FOR SEQ ID NO: 17:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 8 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: unknown  
 TOPOLOGY: unknown  
 MOLECULE TYPE: peptide

US-08-197-484-17

Query Match 0.8%; Score 6; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred.No. 3e+05;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPDEL 736

Db 1 KPDEL 6  
 |||||



; FILING DATE: 29-JAN-1992  
 ; PRIOR APPLICATION DATA: US 07/749,568  
 ; APPLICATION NUMBER: US 07/749,568  
 ; FILING DATE: 26-AUG-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Parmelee, Steven W.  
 ; REGISTRATION NUMBER: 31,990  
 ; REFERENCE/DOCKET NUMBER: 14137-26-4PC  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (206) 467-9500  
 ; TELEFAX: (415) 543-5043  
 ; INFORMATION FOR SEQ ID NO: 17:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 8 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: unknown  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: peptide  
 ; PCT-US95-02121-17

Query Match 0.8%; Score 6; DB 5; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 3e+05;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 731 KPKDEL 736  
 Db 1 KPKDEL 6

RESULT 5  
 US-09-187-859-2953  
 ; Sequence 2953, Application US/09187859A  
 ; Patent No. 6358920  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Blaschuk, Orest W.  
 ; APPLICANT: Gour, Barbara J.  
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL  
 ; TITLE OF INVENTION: CADHERIN-MEDIATED FUNCTIONS  
 ; FILE REFERENCE: 100086.407C1  
 ; CURRENT APPLICATION NUMBER: US/09/187.859A  
 ; CURRENT FILING DATE: 1998-11-06  
 ; NUMBER OF SEQ ID NOS: 4052  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 2953  
 ; LENGTH: 5  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Representative cyclic modulating agent based on  
 ; OTHER INFORMATION: Protocadherin cell adhesion recognition sequence  
 US-09-187-859-2953

Query Match 0.7%; Score 5; DB 4; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 3e+05;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 691 DDLVK 695  
 Db 1 DDLVK 5

RESULT 6  
 US-09-839-542B-2953  
 ; Sequence 2953, Application US/09839542B  
 ; Patent No. 6569996  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Blaschuk, Orest W.  
 ; APPLICANT: Symonds, James Matthew  
 ; APPLICANT: Gour, Barbara J.  
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL  
 ; TITLE OF INVENTION: CADHERIN-MEDIATED FUNCTIONS  
 ; FILE REFERENCE: 100086.407D1  
 ; CURRENT APPLICATION NUMBER: US/09/839,542B

; CURRENT FILING DATE: 2001-04-20  
 ; NUMBER OF SEQ ID NOS: 4052  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 2953  
 ; LENGTH: 5  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Representative cyclic modulating agent based on  
 ; OTHER INFORMATION: Protocadherin cell adhesion recognition sequence  
 US-09-839-542B-2953

Query Match 0.7%; Score 5; DB 4; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 3e+05;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 691 DDLVK 695  
 Db 1 DDLVK 5

RESULT 7  
 US-09-535-852-916  
 ; Sequence 916, Application US/09535852  
 ; Patent No. 6638911  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Blaschuk, Orest W.  
 ; APPLICANT: Symonds, James M.  
 ; APPLICANT: Gour, Barbara J.  
 ; TITLE OF INVENTION: COMPOUNDS AND ETHODS FOR MODULATING  
 ; TITLE OF INVENTION: DESMOSOMAL CADHERIN-MEDIATED FUNCTIONS  
 ; FILE REFERENCE: 100086.407C6  
 ; CURRENT APPLICATION NUMBER: US/09/535,852  
 ; CURRENT FILING DATE: 2001-05-21  
 ; NUMBER OF SEQ ID NOS: 2009  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 916  
 ; LENGTH: 5  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Cyclicized modulating agent comprising  
 ; OTHER INFORMATION: desmoglein-1 cell adhesion recognition sequence  
 US-09-535-852-916

Query Match 0.7%; Score 5; DB 4; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 3e+05;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 242 CRALC 246  
 Db 1 CRALC 5

RESULT 8  
 US-07-609-716-7  
 ; Sequence 7, Application US/07609716  
 ; Patent No. 5514581  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ferrari, Franco A.  
 ; APPLICANT: Cappello, Joseph  
 ; TITLE OF INVENTION: Functional Recombinantly Prepared  
 ; TITLE OF INVENTION: Synthetic Protein Polymer  
 ; NUMBER OF SEQUENCES: 118  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
 ; STREET: Four Embarcadero Center, Suite 3400  
 ; CITY: San Francisco  
 ; STATE: CA  
 ; COUNTRY: US  
 ; ZIP: 94111  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/609,716
; FILING DATE: 06-NOV-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I.
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-55186-3/BIR
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-609-716-7

Query Match      0.7%; Score 5; DB 1; Length 6;
Best Local Similarity 100.0%; Pred.No. 3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      99 SGAGA 103
DB      1 SGAGA 5

RESULT 9
US-08-175-155-1
; Sequence 1, Application US/08175155
; Patent No. 5641648
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; APPLICANT: Criseman, John W.
; APPLICANT: Dorman, Mary A.
; TITLE OF INVENTION: Methods for Preparing Synthetic
; TITLE OF INVENTION: Repetitive DNA
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hombach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/175,155
; FILING DATE: 29-DEC-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I.
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-55186-5/BIR
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: peptide
US-08-175-155-1

Query Match      0.7%; Score 5; DB 1; Length 6;
Best Local Similarity 100.0%; Pred.No. 3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      99 SGAGA 103
DB      1 SGAGA 5

RESULT 10
US-08-469-005A-5
; Sequence 5, Application US/08469005A
; Patent No. 5665874
; GENERAL INFORMATION:
; APPLICANT: KUHAJDA, FRANCIS P.
; APPLICANT: PASTERNAK, GARY A.
; TITLE OF INVENTION: CANCER RELATED ANTIGEN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BAKER & BOTTS, L.L.P.
; STREET: 1299 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004-2400
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,005A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/188,426
; FILING DATE: 24-JAN-1994
; APPLICATION NUMBER: 08/096,908
; FILING DATE: 26-JUL-1993
; APPLICATION NUMBER: 07/917,716
; FILING DATE: 24-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Posorske, Laurence H
; REGISTRATION NUMBER: 34,698
; REFERENCE/DOCKET NUMBER: 062482-0113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-639-7700
; TELEFAX: 202-639-7890
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; US-08-469-005A-5

Query Match      0.7%; Score 5; DB 1; Length 6;
Best Local Similarity 100.0%; Pred.No. 3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      230 AALQE 234
DB      2 AALQE 6

```

```

RESULT 11
US-08-211-747-4
; Sequence 4, Application US/08211747
; Patent No. 5733549
; GENERAL INFORMATION:
; APPLICANT: Yamada, Shingo
; APPLICANT: Inoue, Keiichi
; APPLICANT: Kitajima, Megumi
; APPLICANT: Yoshimura, Hajime
; APPLICANT: Sakurabayashi, Ikunosuke
; TITLE OF INVENTION: Peptides Including Amino Acid Sequences
; TITLE OF INVENTION: Selected From Lipoprotein (a) and Apolipoprotein (a),
; TITLE OF INVENTION: Antibodies Recognizing These Amino Acid Sequences, and
; TITLE OF INVENTION: Methods of Determination Using Antibodies
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10039-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/211.747
; FILING DATE: 14-APR-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7005-077
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; MOLECULE TYPE: peptide
; US-08-211-747-4

Query Match 0.7%; Score 5; DB 1; Length 6;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 AEGTA 44
Db 2 AEGTA 6

RESULT 12
US-08-188-426-5
; Sequence 5, Application US/08188426
; Patent No. 5759791
; GENERAL INFORMATION:
; APPLICANT: Kuhajda, Francis P.
; APPLICANT: Pasternack, Gary R.
; TITLE OF INVENTION: CANCER RELATED ANTIGEN
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, N.W. -- 11th Floor
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.

```

```

; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188.426
; FILING DATE: 24-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Posorske, Laurence H.
; REGISTRATION NUMBER: 34,698
; REFERENCE/DOCKET NUMBER: 1107.45253
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 508-9100
; TELEFAX: 202 508-9299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-188-426-5

Query Match 0.7%; Score 5; DB 1; Length 6;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 AALQE 234
Db 2 AALQE 6

RESULT 13
US-08-477-509B-21
; Sequence 21, Application US/08477509B
; Patent No. 5770697
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A
; APPLICANT: Cappello, Joseph
; APPLICANT: Crissman, John W
; APPLICANT: Dorman, Mary A
; TITLE OF INVENTION: No. 5770697el Peptides Comprising Repetitive
; TITLE OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same
; NUMBER OF SEQUENCES: 112
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477.509B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,155
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/053,049
; FILING DATE: 22-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618

```

;; FILING DATE: 29-OCT-1987  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 06/927,258  
;; FILING DATE: 04-NOV-1986  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Trecartin, Richard F.  
;; REGISTRATION NUMBER: 31,801  
;; REFERENCE/DOCKET NUMBER: A-55186-7/RFT/MTK  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415-781-1989  
;; TELEX: 910 277299  
;; INFORMATION FOR SEQ ID NO: 21:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 6 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
US-08-477-509B-21

Query Match 0.7%; Score 5; DB 1; Length 6;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 SGAGA 103  
|||||  
Db 1 SGAGA 5

RESULT 14  
US-08-707-237A-6  
;; Sequence 6, Application US/08707237A  
;; Patent No. 5820713  
;; GENERAL INFORMATION:  
;; APPLICANT: Ferrari, Franco A.  
;; APPLICANT: Capello, Joseph  
;; APPLICANT: Criseman, John W.  
;; APPLICANT: Dorman, Mary A.  
;; TITLE OF INVENTION: METHODS FOR PREPARING SYNTHETIC  
;; TITLE OF INVENTION: REPETITIVE DNA  
;; NUMBER OF SEQUENCES: 108  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Flehr, Holbach, Test, Albritton & Herbert  
;; STREET: Four Embarcadero Center, Suite 3400  
;; CITY: San Francisco  
;; STATE: California  
;; COUNTRY: United States  
;; ZIP: 94111-4187  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/707,237A  
;; FILING DATE: 03-SEP-1996  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/175,155  
;; FILING DATE: 29-DEC-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/053,049  
;; FILING DATE: 22-APR-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/609,716  
;; FILING DATE: 06-NOV-1990  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/269,429  
;; FILING DATE: 09-NOV-1988  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/114,618  
;; FILING DATE: 29-OCT-1987  
;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 06/927,258  
;; FILING DATE: 04-NOV-1986  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Trecartin, Richard F.  
;; REGISTRATION NUMBER: 31,801  
;; REFERENCE/DOCKET NUMBER: A-55186-10/WHD  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 781-1989  
;; TELEX: 910 277299  
;; INFORMATION FOR SEQ ID NO: 6:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 6 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: unknown  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: protein  
US-08-707-237A-6

Query Match 0.7%; Score 5; DB 2; Length 6;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 SGAGA 103  
|||||  
Db 1 SGAGA 5

RESULT 15  
US-08-659-486A-4  
;; Sequence 4, Application US/08659486A  
;; Patent No. 5840499  
;; GENERAL INFORMATION:  
;; APPLICANT: Lawrence Brass and James A. Hoxie  
;; TITLE OF INVENTION: Method and Kit for the Detection of  
;; TITLE OF INVENTION: Thrombin Receptor Activation of Platelet and other Cells  
;; NUMBER OF SEQUENCES: 4  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Jane Massey Licata, Esq.  
;; STREET: 210 Lake Drive East, Suite 201  
;; CITY: Cherry Hill  
;; STATE: NJ  
;; COUNTRY: USA  
;; ZIP: 08002  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
;; COMPUTER: IBM 486  
;; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS  
;; SOFTWARE: WORDPERFECT 5.1  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/659,486A  
;; FILING DATE: June 6, 1996  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/220,375  
;; FILING DATE: March 31, 1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Jane Massey Licata  
;; REGISTRATION NUMBER: 32,257  
;; REFERENCE/DOCKET NUMBER: PENN-0331  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (609) 779-2400  
;; TELEFAX: (609) 779-8488  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 6  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
US-08-659-486A-4

Query Match 0.7%; Score 5; DB 2; Length 6;  
Best Local Similarity 100.0%; Pred. No. 3e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 677 TLDPR 681  
| | | | |  
Db 2 TLDPR 6

Search completed: August 17, 2004, 10:17:38  
Job time : 20 secs

This Page Blank (uspio)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2004, 09:45:24 ; Search time 24 Seconds  
(without alignments)  
2949.873 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 736  
Sequence: 1 MAVRALKLLTTLAVVAAS.....PPEPAQESLSGSESKPKDEL 736

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 293366 seqs, 96191526 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.2	364	1 A55014	GTP-binding protei
2	8	1.1	222	2 H81354	probable undecap
3	8	1.1	257	2 T32002	hypothetical prote
4	8	1.1	317	2 T35010	probable integral
5	8	1.1	431	2 A56922	synaptonemal compl
6	8	1.1	600	2 T41165	hypothetical prote
7	8	1.1	623	2 AB1112	transcription anti
8	8	1.1	623	2 AF1473	transcription anti
9	8	1.1	913	2 T51923	related to heterok
10	8	1.1	973	2 T01862	hypothetical prote
11	8	1.1	1233	1 G71612	novel serine/threo
12	8	1.1	1384	2 T02748	hypothetical prote
13	8	1.1	2895	2 H85362	hypothetical prote
14	7	1.0	34	2 F82163	hypothetical prote
15	7	1.0	54	2 C84292	hypothetical prote
16	7	1.0	73	2 AC0914	probable lipoprote
17	7	1.0	76	2 B71840	hypothetical prote
18	7	1.0	76	2 B64675	conserved hypotet
19	7	1.0	81	2 T03711	reverse transcript
20	7	1.0	82	2 A61040	transforming growt
21	7	1.0	89	2 S72742	B117 Fl 32 protei
22	7	1.0	103	2 S37966	hypothetical prote
23	7	1.0	107	2 T23910	hypothetical prote
24	7	1.0	110	2 B70814	hypothetical prote
25	7	1.0	121	2 B58134	transforming growt
26	7	1.0	123	2 B30560	Ig heavy chain V r
27	7	1.0	129	2 AD0029	50S ribosomal prot
28	7	1.0	140	2 G69813	transcription regu
29	7	1.0	142	2 E95176	acetyltransferase,

30 7 1.0 145 2 C87674  
31 7 1.0 157 2 H83543  
32 7 1.0 159 1 WPT1  
33 7 1.0 159 1 S27195  
34 7 1.0 159 2 IS7497  
35 7 1.0 160 1 WTHU1  
36 7 1.0 160 2 S39795  
37 7 1.0 160 2 G98042  
38 7 1.0 171 2 B75446  
39 7 1.0 177 2 F76764  
40 7 1.0 177 2 AE2899  
41 7 1.0 180 2 D90318  
42 7 1.0 188 2 I79500  
43 7 1.0 188 2 JU0451  
44 7 1.0 188 2 I59116  
45 7 1.0 188 2 A72745

#### ALIGNMENTS

##### RESULT 1

A55014  
GTP-binding protein DRG homolog - human  
C:Species: Homo sapiens (man)  
C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 19-Jan-2001  
R:Schenker, T.; Lach, C.; Kessler, B.; Calderara, S.; Trueb, B.  
J. Biol. Chem. 269, 25447-25453, 1994  
A:Title: A novel GTP-binding protein which is selectively repressed in SV40 transformed f  
A:Reference number: A55014; MUID:95014343; PMID:7929244  
A:Accession: A55014  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-364 <SCH>  
A:Cross-references: GB:X80754; NID:g577778; PIDN:CAA56730.1; PID:g577779  
A:Experimental source: fibroblasts  
C:Comment: This protein is expressed in normal fibroblasts but repressed in transformed c  
C:Genetics:  
A:Gene: GDB:NEDD3; DRG1  
A:Cross-references: GDB:270141  
A:Superfamily: GTP-binding protein DRG; translation elongation factor Tu homology  
C:Keywords: GTP binding; nucleotide binding; P-loop  
F:63-174/Domain: translation elongation factor Tu homology <ETU>  
F:69-76/Region: nucleotide-binding motif A (P-loop)  
F:92-97/Region: GTP binding #status predicted  
F:115-118/Region: GTP binding #status predicted  
F:246-249/Region: GTP binding #status predicted  
F:337-341/Region: GTP binding #status predicted

Query Match

Best Local Similarity 1.2%; Score 9; DB 1; Length 364;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 367 QRSLLLEKEL 375  
DB 155 QRSLLLEKEL 163

##### RESULT 2

H81354  
probable undecaprenyl diphosphate synthase Cj0824 [imported] - Campylobacter jejuni (str  
C:Species: Campylobacter jejuni  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
C:Accession: H81354  
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kettle, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whithead, S.; Barrell  
Nature 403, 665-668, 2000  
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hype  
A:Reference number: A81250; MUID:20150912; PMID:10688204  
A:Accession: H81354  
A>Status: preliminary  
A:Molecule type: DNA

A;Residues: 1-222 <PAR>  
A;Cross-references: GB:AL119076; GB:AL111168; NID:G6968128; PIDN:CAE73089.1; PID:G696827  
A;Experimental source: serotype O2, strain NCTC 11168  
C;Genetics:  
A;Gene: upps; Cj0824  
C;Superfamily: conserved hypothetical protein YER002c

Query Match 1.1%; Score 8; DB 2; Length 222;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 430 TLVEKTK 437  
|||  
DB 111 TLVEKTK 118

RESULT 3  
T32002  
hypothetical protein F36H9.3 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 31-Jan-2000  
C;Accession: T32002  
R;Daute, M.; Kramer, J.; Twyman, B.  
submitted to the EMBL Data Library, July 1997  
A;Description: The sequence of C. elegans cosmid F36H9.  
A;Reference number: Z2110  
A;Accession: T32002  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-257 <DAN>  
A;Cross-references: EMBL:AF016668; PIDN:AAB66092.1; GSPDB:GN00020; CESP:F36H9.3  
A;Experimental source: strain Bristol N2; clone F36H9  
C;Genetics:  
A;Gene: CESP:F36H9.3  
A;Map position: 2  
A;Introns: 21/3; 49/3; 192/1  
C;Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 1.1%; Score 8; DB 2; Length 257;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 EAVPHLEA 230  
|||  
DB 130 EAVPHLEA 137

RESULT 4  
T35010  
probable integral membrane protein - Streptomyces coelicolor  
C;Species: Streptomyces coelicolor  
C;Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
C;Accession: T35010  
R;Seeger, S.; Harris, D.; Jones, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, June 1999  
A;Reference number: Z21565  
A;Accession: T35010  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-317 <SEE>  
A;Cross-references: EMBL:AL079355; PIDN:CAB45566.1; GSPDB:GN00070; SCOEDB:SC4C6.02c  
A;Experimental source: strain A3(2)  
C;Genetics:  
A;Gene: SCOEDB:SC4C6.02c

Query Match 1.1%; Score 8; DB 2; Length 317;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 ASGAGALR 105  
|||  
DB 32 ASGAGALR 39

RESULT 5  
A56822  
synaptonemal complex protein SC56 - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 05-Jan-1996 #sequence\_revision 05-Jan-1996 #text\_change 04-Mar-2000  
C;Accession: A56822; S20742  
R;Chen, Q.; Pearlman, R.E.; Moens, P.B.  
Biochem. Cell Biol. 70, 1030-1038, 1992  
A;Title: Isolation and characterization of a cDNA encoding a synaptonemal complex protein  
A;Reference number: A56822; MUID:93213429; PMID:1363622  
A;Accession: A56822  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-431 <CHE>  
A;Cross-references: EMBL:X65454; NID:G57191; PIDN:CRA46449.1; PID:G57192  
A;Experimental source: testis  
A;Note: sequence extracted from NCBI backbone (NCBIP:129004)  
C;Superfamily: rat synaptonemal complex protein SC56

Query Match 1.1%; Score 8; DB 2; Length 431;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 RAACLRRRC 123  
|||  
DB 85 RAACLRRRC 92

RESULT 6  
T41165  
hypothetical protein SPCC1827.04 - fission yeast (Schizosaccharomyces pombe)  
C;Species: Schizosaccharomyces pombe  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 28-Jul-2000  
C;Accession: T41165  
R;Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, June 1998  
A;Reference number: Z21974  
A;Accession: T41165  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-600 <SEE>  
A;Cross-references: EMBL:AL023777; PIDN:CAA19312.1; GSPDB:GN00068; SPDB:SPCC1827.04  
A;Experimental source: strain 972h-; cosmid cl827  
C;Genetics:  
A;Gene: SPDB:SPCC1827.04  
A;Map position: 3  
C;Superfamily: Schizosaccharomyces pombe hypothetical protein SPCC1827.04

Query Match 1.1%; Score 8; DB 2; Length 600;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 723 ESLSGSES 730  
|||  
DB 105 ESLSGSES 112

RESULT 7  
AB1112  
transcription antiterminator (BglG family) homolog lmo0297 [imported] - Listeria monocytogenes  
C;Species: Listeria monocytogenes  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C;Accession: AB1112  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkut, G.; Madueno, E.; Maitournam, A.; Matok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AB1112



A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-623 <GLA>  
A;Cross-references: GB:NC\_003210; PIDN:CAD00824.1; PID:g16409662; GSPDB:GN00177  
A;Experimental source: strain EGD-e  
C;Genetics:  
A;Gene: lmo0297

Query Match 1.1%; Score 8; DB 2; Length 623;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 366 RORSLEK 373  
Db 6 RORSLEK 13  
|||||

RESULT 8  
AF1473  
transcription antiterminator (BglG family) homolog lin0325 [imported] - *Listeria innocua*  
C;Species: *Listeria innocua*  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C;Accession: AF1473  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H. D.; Jones, L.M.; Karst, U. Science 294, 849-852, 2001  
A;Authors: Krefft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mak, C.; Schluster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria* species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AF1473  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-623 <GLA>  
A;Cross-references: GB:AL592022; PIDN:CAC95558.1; PID:g16412754; GSPDB:GN00178  
A;Experimental source: strain clipl1262  
C;Genetics:  
A;Gene: lin0325

Query Match 1.1%; Score 8; DB 2; Length 623;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 366 RORSLEK 373  
Db 6 RORSLEK 13  
|||||

RESULT 9  
T51923  
related to heterokaryon incompatibility protein het-6 [imported] - *Neurospora crassa*  
N;Alternate names: protein B23111.350  
C;Species: *Neurospora crassa*  
C;Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 20-Oct-2000  
C;Accession: T51923  
R;Schulze, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, August 2000  
A;Reference number: 225958  
A;Accession: T51923  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-913 <SCH>  
A;Cross-references: EMBL:AL391572; GSPDB:GN00116; NCSP:B23111.350  
A;Experimental source: BAC clone B23111; strain OR74A  
C;Genetics:  
A;Gene: NCSP:B23111.350  
A;Map position: 6  
A;Introns: 13/1; 410/2; 516/2; 867/3; 887/3

Query Match 1.1%; Score 8; DB 2; Length 913;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 66 AALRALRL 73  
Db 213 AALRALRL 220  
|||||

RESULT 10  
T01862  
hypothetical protein T7M24.4 - *Arabidopsis thaliana*  
C;Species: *Arabidopsis thaliana* (mouse-ear cress)  
C;Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 24-Mar-1999  
C;Accession: T01862  
R;Harmon, G.; Langston, Y.; Stoneking, T.; Drone, K.; Ames, M. submitted to the EMBL Data Library, July 1998  
A;Description: The sequence of *Arabidopsis thaliana* T7M24.  
A;Reference number: Z14448  
A;Accession: T01862  
A;Status: translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-973 <HAR>  
A;Cross-references: EMBL:AF077408; NID:g3319359; PID:g3319363  
A;Experimental source: cultivar Columbia  
C;Genetics:  
A;Map position: 4  
A;Introns: 257/1; 284/1; 378/3; 689/3; 744/3  
A;Note: T7M24.4

Query Match 1.1%; Score 8; DB 2; Length 973;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 429 ETLVBKKT 436  
Db 19 ETLVBKKT 26  
|||||

RESULT 11  
G71612  
novel serine/threonine-specific protein kinase (EC 2.7.1.-) PF0520W - malaria parasite  
C;Species: *Plasmodium falciparum*  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jun-2000  
C;Accession: G71612  
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.; Science 282, 1126-1132, 1998  
A;Title: Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*.  
A;Reference number: AF1600; MUID:99021743; PMID:9804551  
A;Accession: G71612  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-1233 <GAR>  
A;Cross-references: GB:AE001400; GB:AE001362; NID:g3845204; PIDN:AACT1893.1; PID:g3845206  
A;Experimental source: clone 3D7  
C;Genetics:  
A;Gene: PF0520W  
C;Superfamily: malaria parasite serine/threonine-specific protein kinase PF0520W; protein F; 949-1221/Domain: protein kinase homology <KIN>

Query Match 1.1%; Score 8; DB 1; Length 1233;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 154 FKINKLEK 161  
Db 771 FKINKLEK 778  
|||||

RESULT 12  
T02748  
hypothetical protein R31180.1 - human  
C;Species: *Homo sapiens* (man)  
C;Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 31-Dec-2000

C:Accession: T02748  
 R:lamerdin, J.E.; McCready, P.M.; Skowronski, E.; Adamson, A.W.; Burkhardt-Schultz, K.; G  
 J.; Danganan, L.; Poundstone, P.; Christensen, M.; Georgescu, A.; Avila, J.; Liu, S.; Ab  
 P.; Quan, G.; Krommiller, B.; Arellano, A.; Montgomery, M.; Ow, D.; Nolan, M.  
 Submitted to the EMBL Data Library, August 1998  
 A:Authors: Trong, S.; Kobayashi, A.; Olsen, A.S.; Carrano, A.V.  
 A:Description: Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a serine  
 A:Reference number: Z14714  
 A:Accession: T02748  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1384 <LAM>  
 A:Cross-references: EMBL:AC005390; NID:g33399675; PIDN:AAC28919.1; PID:g33399676  
 C:Genetics:  
 A:Map position: 19  
 A:Introns: 15/3; 51/1; 131/2; 180/3; 216/1; 266/3; 311/2; 359/3; 407/3; 447/1; 482/2; 53  
 3; 1160/2; 1224/1  
 C:Superfamily: Arabidopsis thaliana hypothetical protein YUP8H12R.3

Query Match 1.1%; Score 8; DB 2; Length 1384;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 296 FEDFLPSH 303  
 |||||  
 DB 164 FEDFLPSH 171  
 |||||

RESULT 13  
 H85362  
 hypothetical protein AT4G30990 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
 C:Accession: H85362  
 R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
 Nature 402, 769-777, 1999  
 A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
 A:Reference number: A85001; MUID:20083488; PMID:10617198  
 A:Accession: H85362  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2895 <STO>  
 A:Cross-references: GB:NC\_001269; NID:g2720001; PIDN:CAB79817.1; GSPDB:GN00140  
 A:Gene: AT4G30990  
 A:Map position: 4

Query Match 1.1%; Score 8; DB 2; Length 2895;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LKLTLL 13  
 |||||  
 DB 1916 LKLTLL 1923  
 |||||

RESULT 14  
 F82163  
 hypothetical protein VC1734 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
 C:Species: Vibrio cholerae  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
 C:Accession: F82163  
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
 chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, R.  
 L.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406, 477-483, 2000  
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
 A:Reference number: A82035; MUID:20406833; PMID:10952301  
 A:Accession: F82163  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-34 <HEI>  
 A:Cross-references: GB:AE004251; GB:AE003852; NID:g956248; PIDN:AAF94884.1; GSPDB:GN001

A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C:Genetics:  
 A:Gene: VC1734  
 A:Map position: 1

Query Match 1.0%; Score 7; DB 2; Length 34;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 AAHSLSE 134  
 |||||  
 DB 11 AAHSLSE 17  
 |||||

RESULT 15  
 C84292  
 hypothetical protein Vng1376h [imported] - Halobacterium sp. NRC-1  
 C:Species: Halobacterium sp. NRC-1  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: C84292  
 R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;  
 Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon  
 Jung, K.H.; Alam, M.; Freitas, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Lia  
 A:Title: Genome sequence of Halobacterium species NRC-1.  
 A:Reference number: A84160; MUID:20504483; PMID:11016950  
 A:Accession: C84292  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-54 <STO>  
 A:Cross-references: GB:AE004437; NID:gi0580885; PIDN:AAG19703.1; GSPDB:GN00138  
 C:Genetics:  
 A:Gene: VNG1376H

Query Match 1.0%; Score 7; DB 2; Length 54;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 HLEAALQ 233  
 |||||  
 DB 11 HLEAALQ 17  
 |||||

Search completed: July 18, 2004, 10:00:23  
 Job time : 26 secs

Query Match 0.8%; Score 6; DB 12; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels

QY 731 KPKDEL 736  
 |||||  
 DB 1 KPKDEL 6

## RESULT 2

US-10-367-593-129  
 ; Sequence 129, Application US/10367593  
 ; Publication No. US20040071721A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rothman, James E.  
 ; APPLICANT: Hartl, F. Ulrich  
 ; APPLICANT: Hoe, Mee H.  
 ; APPLICANT: Houghton, Alan  
 ; APPLICANT: Takechi, Yoshizumi  
 ; APPLICANT: Mayhew, Mark  
 ; TITLE OF INVENTION: Heat Shock Protein-Based Vaccines and Immunotherapies  
 ; FILE REFERENCE: 11746/461012  
 ; CURRENT APPLICATION NUMBER: US/10/367,593  
 ; CURRENT FILING DATE: 2003-02-14  
 ; PRIOR APPLICATION NUMBER: US 09/011,645  
 ; PRIOR FILING DATE: 1998-02-13  
 ; PRIOR APPLICATION NUMBER: PCT/US96/13363  
 ; PRIOR FILING DATE: 1996-08-16  
 ; PRIOR APPLICATION NUMBER: US 60/002,490  
 ; PRIOR FILING DATE: 1995-08-18  
 ; PRIOR APPLICATION NUMBER: US 60/002,479  
 ; PRIOR FILING DATE: 1995-08-18  
 ; NUMBER OF SEQ ID NOS: 349  
 ; SOFTWARE: WordPerfect 8.0 for Windows  
 ; SEQ ID NO 129  
 ; LENGTH: 8  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: synthetic peptide  
 US-10-367-593-129

Query Match 0.8%; Score 6; DB 12; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPKDEL 736  
 |||||  
 DB 1 KPKDEL 6

## RESULT 3

US-10-367-594-129  
 ; Sequence 129, Application US/10367594  
 ; Publication No. US20040071722A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rothman, James E.  
 ; APPLICANT: Hartl, F. Ulrich  
 ; APPLICANT: Hoe, Mee H.  
 ; APPLICANT: Houghton, Alan  
 ; APPLICANT: Takechi, Yoshizumi  
 ; APPLICANT: Mayhew, Mark  
 ; TITLE OF INVENTION: Heat Shock Protein-Based Vaccines and Immunotherapies  
 ; FILE REFERENCE: 11746/461041  
 ; CURRENT APPLICATION NUMBER: US/10/367,594  
 ; CURRENT FILING DATE: 2003-02-14  
 ; PRIOR APPLICATION NUMBER: US 09/680,806  
 ; PRIOR FILING DATE: 2000-10-05  
 ; PRIOR APPLICATION NUMBER: US 09/011,645  
 ; PRIOR FILING DATE: 1998-02-13  
 ; PRIOR APPLICATION NUMBER: PCT/US96/13363  
 ; PRIOR FILING DATE: 1996-08-16  
 ; PRIOR APPLICATION NUMBER: US 60/002,490  
 ; PRIOR FILING DATE: 1995-08-18  
 ; PRIOR APPLICATION NUMBER: US 60/002,479  
 ; PRIOR FILING DATE: 1995-08-18

Query Match 0.8%; Score 6; DB 12; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

; NUMBER OF SEQ ID NOS: 349  
 ; SOFTWARE: WordPerfect 8.0 for Windows  
 ; SEQ ID NO 129  
 ; LENGTH: 8  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: synthetic peptide  
 US-10-367-594-129

Query Match 0.8%; Score 6; DB 12; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPKDEL 736  
 |||||  
 DB 1 KPKDEL 6

## RESULT 4

US-10-367-654-129  
 ; Sequence 129, Application US/10367654  
 ; Publication No. US20040071723A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rothman, James E.  
 ; APPLICANT: Hartl, F. Ulrich  
 ; APPLICANT: Hoe, Mee H.  
 ; APPLICANT: Houghton, Alan  
 ; APPLICANT: Takechi, Yoshizumi  
 ; APPLICANT: Mayhew, Mark  
 ; TITLE OF INVENTION: Heat Shock Protein-Based Vaccines and Immunotherapies  
 ; FILE REFERENCE: 11746/461032  
 ; CURRENT APPLICATION NUMBER: US/10/367,654  
 ; CURRENT FILING DATE: 2003-02-14  
 ; PRIOR APPLICATION NUMBER: US 10/171,734  
 ; PRIOR FILING DATE: 2002-06-13  
 ; PRIOR APPLICATION NUMBER: US 09/636,295  
 ; PRIOR FILING DATE: 2000-08-10  
 ; PRIOR APPLICATION NUMBER: US 09/011,645  
 ; PRIOR FILING DATE: 1998-02-13  
 ; PRIOR APPLICATION NUMBER: PCT/US96/13363  
 ; PRIOR FILING DATE: 1996-08-16  
 ; PRIOR APPLICATION NUMBER: US 60/002,490  
 ; PRIOR FILING DATE: 1995-08-18  
 ; PRIOR APPLICATION NUMBER: US 60/002,479  
 ; PRIOR FILING DATE: 1995-08-18  
 ; NUMBER OF SEQ ID NOS: 349  
 ; SOFTWARE: WordPerfect 8.0 for Windows  
 ; SEQ ID NO 129  
 ; LENGTH: 8  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: synthetic peptide  
 US-10-367-654-129

Query Match 0.8%; Score 6; DB 12; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPKDEL 736  
 |||||  
 DB 1 KPKDEL 6

## RESULT 5

US-10-367-658-129  
 ; Sequence 129, Application US/10367658  
 ; Publication No. US20040071724A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rothman, James E.  
 ; APPLICANT: Hartl, F. Ulrich  
 ; APPLICANT: Hoe, Mee H.

APPLICANT: Houghton, Alan  
APPLICANT: Takechi, Yoshizumi  
APPLICANT: Mayhew, Mark  
TITLE OF INVENTION: Heat Shock Protein-Based Vaccines and Immunotherapies  
FILE REFERENCE: 11746/461051  
CURRENT APPLICATION NUMBER: US/10/367,658  
CURRENT FILING DATE: 2003-02-14  
PRIOR APPLICATION NUMBER: US 09/794,529  
PRIOR FILING DATE: 2001-02-27  
PRIOR APPLICATION NUMBER: US 09/011,645  
PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: PCT/US96/13363  
PRIOR FILING DATE: 1996-08-16  
PRIOR APPLICATION NUMBER: US 60/002,490  
PRIOR FILING DATE: 1995-08-18  
PRIOR APPLICATION NUMBER: US 60/002,479  
PRIOR FILING DATE: 1995-08-18  
NUMBER OF SEQ ID NOS: 349  
SOFTWARE: WordPerfect 8.0 for Windows  
SEQ ID NO 129  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: synthetic peptide  
US-10-367-658-129

Query Match 0.8%; Score 6; DB 12; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 731 KPDEL 736  
Db 1 KPDEL 6

RESULT 6  
US-10-367-668-129  
Sequence 129, Application US/10367668  
Publication No. US20040071725A1  
GENERAL INFORMATION:  
APPLICANT: Rothman, James E.  
APPLICANT: Hartl, F. Ulrich  
APPLICANT: Hoe, Mee H.  
APPLICANT: Houghton, Alan  
APPLICANT: Takechi, Yoshizumi  
APPLICANT: Mayhew, Mark  
TITLE OF INVENTION: Heat Shock Protein-Based Vaccines and Immunotherapies  
FILE REFERENCE: 11746/461072  
CURRENT APPLICATION NUMBER: US/10/367,668  
CURRENT FILING DATE: 2003-02-14  
PRIOR APPLICATION NUMBER: US 09/794,517  
PRIOR FILING DATE: 2001-02-27  
PRIOR APPLICATION NUMBER: US 09/011,645  
PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: PCT/US96/13363  
PRIOR FILING DATE: 1996-08-16  
PRIOR APPLICATION NUMBER: US 60/002,490  
PRIOR FILING DATE: 1995-08-18  
PRIOR APPLICATION NUMBER: US 60/002,479  
PRIOR FILING DATE: 1995-08-18  
NUMBER OF SEQ ID NOS: 349  
SOFTWARE: WordPerfect 8.0 for Windows  
SEQ ID NO 129  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: synthetic peptide  
US-10-367-668-129

Query Match 0.8%; Score 6; DB 12; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 731 KPDEL 736  
Db 1 KPDEL 6

RESULT 7  
US-10-128-711-17  
Sequence 17, Application US/10128711  
Publication No. US20030099634A1  
GENERAL INFORMATION:  
APPLICANT: VITIELLO, Maria A.  
APPLICANT: CHESTNUT, Robert W.  
APPLICANT: SETTE, Alessandro D.  
APPLICANT: CELIS, Esteban  
APPLICANT: GRAY, Howard  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR ELICITING CTL IMMUNITY  
NUMBER OF SEQUENCES: 153  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend Kourie and Crew  
STREET: Steuart Street Tower, One Market Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: US  
ZIP: 94105-1493  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/128,711  
FILING DATE: 22-Apr-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/197,484  
FILING DATE: 16-FEB-1994  
APPLICATION NUMBER: US 07/935,811  
FILING DATE: 26-AUG-1992  
APPLICATION NUMBER: US 07/874,491  
FILING DATE: 27-APR-1992  
APPLICATION NUMBER: US 07/827,682  
FILING DATE: 29-JAN-1992  
APPLICATION NUMBER: US 07/749,568  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Parmelee, Steven W.  
REGISTRATION NUMBER: 31,990  
REFERENCE/DOCKET NUMBER: 14137-26-4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 467-9600  
TELEFAX: (206) 623-6793  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 17:  
US-10-128-711-17

Query Match 0.8%; Score 6; DB 14; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 731 KPDEL 736  
Db 1 KPDEL 6

## RESULT 8

US-10-079-167-4  
 ; Sequence 4, Application US/10079167  
 ; Publication No. US20030138454A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hill, Adrian V.S.  
 ; APPLICANT: McShane, Helen  
 ; APPLICANT: Gilbert, Sarah C.  
 ; APPLICANT: Reece, William  
 ; APPLICANT: Schneider, Joerg  
 ; TITLE OF INVENTION: Vaccination Method  
 ; FILE REFERENCE: 2907.1000-001  
 ; CURRENT APPLICATION NUMBER: US/10/079,167  
 ; CURRENT FILING DATE: 2002-02-19  
 ; PRIOR APPLICATION NUMBER: US 09/454,204  
 ; PRIOR FILING DATE: 1999-12-09  
 ; PRIOR APPLICATION NUMBER: PCT/GB98/01681  
 ; PRIOR FILING DATE: 1998-06-09  
 ; PRIOR APPLICATION NUMBER: GB 97 11957.2  
 ; PRIOR FILING DATE: 1997-06-09  
 ; PRIOR APPLICATION NUMBER: PCT/GB01/04116  
 ; PRIOR FILING DATE: 2001-09-13  
 ; PRIOR APPLICATION NUMBER: GB 00 23203.3  
 ; PRIOR FILING DATE: 2001-09-21  
 ; NUMBER OF SEQ ID NOS: 99  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 8  
 ; TYPE: PRT  
 ; ORGANISM: Unknown  
 ; FEATURE:  
 ; OTHER INFORMATION: CTL Epitope of the Malaria String  
 US-10-079-167-4

Query Match 0.8%; Score 6; DB 14; Length 8;  
 Best Local Similarity 100.0%; Pred.No. 1.2e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPKDEL 736  
 Db 1 KPKDEL 6

## RESULT 9

US-10-360-836-69  
 ; Sequence 69, Application US/10360836  
 ; Publication No. US20030185854A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zavala, Fidel  
 ; APPLICANT: Birkett, Ashley  
 ; TITLE OF INVENTION: USE OF RECOMBINANT HEPATITIS B CORE  
 ; TITLE OF INVENTION: PARTICLES TO DEVELOP VACCINES AGAINST INFECTIOUS PATHOGENS  
 ; TITLE OF INVENTION: AND MALIGNANCIES  
 ; FILE REFERENCE: 5986/1J876  
 ; CURRENT APPLICATION NUMBER: US/10/360,836  
 ; CURRENT FILING DATE: 2003-02-07  
 ; PRIOR APPLICATION NUMBER: 60/354,963  
 ; PRIOR FILING DATE: 2002-02-08  
 ; NUMBER OF SEQ ID NOS: 86  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 69  
 ; LENGTH: 8  
 ; TYPE: PRT  
 ; ORGANISM: Plasmodium  
 US-10-360-836-69

Query Match 0.8%; Score 6; DB 14; Length 8;  
 Best Local Similarity 100.0%; Pred.No. 1.2e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPKDEL 736  
 Db 1 KPKDEL 6

## RESULT 10

US-10-367-674-129  
 ; Sequence 129, Application US/10367674  
 ; Publication No. US20040127684A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rothman, James E.  
 ; APPLICANT: Hartl, F. Ulrich  
 ; APPLICANT: Hoe, Mee H.  
 ; APPLICANT: Houghton, Alan  
 ; APPLICANT: Takechi, Yoshizumi  
 ; APPLICANT: Mayhew, Mark  
 ; TITLE OF INVENTION: Heat Shock Protein-Based Vaccines and Immunotherapies  
 ; FILE REFERENCE: 11746/4610211  
 ; CURRENT APPLICATION NUMBER: US/10/367,674  
 ; CURRENT FILING DATE: 2003-02-14  
 ; PRIOR APPLICATION NUMBER: US 10/170,738  
 ; PRIOR FILING DATE: 2002-06-13  
 ; PRIOR APPLICATION NUMBER: US 09/552,868  
 ; PRIOR FILING DATE: 2000-04-20  
 ; PRIOR APPLICATION NUMBER: US 09/011,645  
 ; PRIOR FILING DATE: 1998-02-13  
 ; PRIOR APPLICATION NUMBER: PCT/US96/13363  
 ; PRIOR FILING DATE: 1996-08-16  
 ; PRIOR APPLICATION NUMBER: US 60/002,490  
 ; PRIOR FILING DATE: 1995-08-18  
 ; PRIOR APPLICATION NUMBER: US 60/002,479  
 ; PRIOR FILING DATE: 1995-08-18  
 ; NUMBER OF SEQ ID NOS: 349  
 ; SOFTWARE: WordPerfect 8.0 for Windows  
 ; SEQ ID NO 129  
 ; LENGTH: 8  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: synthetic peptide  
 US-10-367-674-129

Query Match 0.8%; Score 6; DB 16; Length 8;  
 Best Local Similarity 100.0%; Pred.No. 1.2e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPKDEL 736  
 Db 1 KPKDEL 6

## RESULT 11

US-10-653-624-4  
 ; Sequence 4, Application US/10653624  
 ; Publication No. US20040131594A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: McMichael, Andrew  
 ; APPLICANT: Hill, Adrian V.S.  
 ; APPLICANT: Gilbert, Sarah C.  
 ; APPLICANT: Schneider, Joerg  
 ; APPLICANT: Plebanski, Magdalena  
 ; APPLICANT: Hanke, Tomas  
 ; APPLICANT: Smith, Geoffrey L.  
 ; APPLICANT: Blanchard, Tom  
 ; TITLE OF INVENTION: Methods and Reagents for Vaccination  
 ; FILE REFERENCE: 2907.1000-000  
 ; CURRENT APPLICATION NUMBER: US/10/653,624  
 ; CURRENT FILING DATE: 2003-09-02  
 ; PRIOR APPLICATION NUMBER: US/09/454,204A  
 ; PRIOR FILING DATE: 1999-12-09  
 ; PRIOR APPLICATION NUMBER: PCT/GB98/01681  
 ; PRIOR FILING DATE: 1998-06-09  
 ; PRIOR APPLICATION NUMBER: GB 97 11957.2  
 ; PRIOR FILING DATE: 1997-06-09  
 ; NUMBER OF SEQ ID NOS: 78

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: CTL Epitope of the Malaria String
US-10-653-624-4

Query Match          0.8%; Score 6; DB 16; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.2e+06;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPKDEL 736
Db 1 KPKDEL 6

RESULT 12
US-10-777-053-768
; Sequence 768, Application US/10777053
; Publication No. US20040132088A1
; GENERAL INFORMATION:
; APPLICANT: Simard, John J. L.
; APPLICANT: Diamond, David C.
; APPLICANT: Qiu, Zhiyong
; APPLICANT: Lei, Xiang-Dong
; TITLE OF INVENTION: EXPRESSION VECTORS ENCODING EPITOPES OF
; TITLE OF INVENTION: TARGET-ASSOCIATED ANTIGENS AND METHODS FOR THEIR DESIGN
; FILE REFERENCE: MANK:022C1
; CURRENT APPLICATION NUMBER: US/10/777,053
; CURRENT FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: 10/292,413
; PRIOR FILING DATE: 2002-11-07
; PRIOR APPLICATION NUMBER: 60/336,968
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 979
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 768
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Plasmodium Falciparum
US-10-777-053-768

Query Match          0.8%; Score 6; DB 16; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.2e+06;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 KPKDEL 736
Db 1 KPKDEL 6

RESULT 13
US-10-006-869-2953
; Sequence 2953, Application US/10006869
; Publication No. US20030082166A1
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Symonds, James Matthew
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
; TITLE OF INVENTION: CADHERIN-MEDIATED FUNCTIONS
; FILE REFERENCE: 100086.407C7
; CURRENT APPLICATION NUMBER: US/10/006,869
; CURRENT FILING DATE: 2001-12-03
; NUMBER OF SEQ ID NOS: 4052
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2953
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: peptide
US-09-800-433-6

Query Match          0.7%; Score 5; DB 9; Length 6;
Best Local Similarity 100.0%; Pred. No. 1.2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 691 DDLVK 695
Db 1 DDLVK 5

RESULT 14
US-10-395-032-2953
; Sequence 2953, Application US/10395032
; Publication No. US20030229199A1
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Symonds, James Matthew
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
; TITLE OF INVENTION: CADHERIN-MEDIATED FUNCTIONS
; FILE REFERENCE: 100086.407C9
; CURRENT APPLICATION NUMBER: US/10/395,032
; CURRENT FILING DATE: 2003-03-21
; NUMBER OF SEQ ID NOS: 4052
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2953
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Representative cyclic modulating agent based on
; OTHER INFORMATION: Protocadherin cell adhesion recognition sequence
US-10-395-032-2953

Query Match          0.7%; Score 5; DB 15; Length 5;
Best Local Similarity 100.0%; Pred. No. 1.2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 691 DDLVK 695
Db 1 DDLVK 5

RESULT 15
US-09-800-433-6
; Sequence 6, Application US/09800433
; Patent No. US20020106378A1
; GENERAL INFORMATION:
; APPLICANT: O'Hare and Elliott
; TITLE OF INVENTION: Materials and methods for intracellular transport and
; TITLE OF INVENTION: their uses
; FILE REFERENCE: 49408
; CURRENT APPLICATION NUMBER: US/09/800,433
; CURRENT FILING DATE: 2000-03-05
; PRIOR APPLICATION NUMBER: 09/395,344
; PRIOR FILING DATE: 1999-09-13
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 6
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: peptide
US-09-800-433-6

Query Match          0.7%; Score 5; DB 9; Length 6;
Best Local Similarity 100.0%; Pred. No. 1.2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 63 RSRAA 67  
|||  
Db 1 RSRAA 5

Search completed: August 17, 2004, 10:18:36  
Job time : 48 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 17, 2004, 10:13:20 ; Search time 20 Seconds  
(without alignments)  
3539.848 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 736  
Sequence: 1 MAVRALKLTLLAVAAAS.....PPEPAQESLSGSESKPKDEL 736

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 605

Minimum DB seq length: 0  
Maximum DB seq length: 8

Post-processing: Listing first 45 summaries

Database : PIR 78.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	5	0.7	6	2 C22565	R-phycoerythrin be
2	5	0.7	6	2 PT0605	T-cell receptor be
3	4	0.5	5	2 PT0703	T-cell receptor be
4	4	0.5	5	2 PT0701	T-cell receptor be
5	4	0.5	6	2 A11490	pyruvate kinase (E
6	4	0.5	6	2 B33932	IG mu chain D regi
7	4	0.5	6	2 PT0604	T-cell receptor be
8	4	0.5	6	2 PT0668	T-cell receptor be
9	4	0.5	7	2 JN0859	peptidyl-di-peptida
10	4	0.5	7	2 PT0654	T-cell receptor be
11	4	0.5	7	2 PT0543	T-cell receptor be
12	4	0.5	7	2 PT0689	T-cell receptor be
13	4	0.5	7	2 PT0719	T-cell receptor be
14	4	0.5	8	2 A32523	peptidyl-di-peptida
15	4	0.5	8	2 JS0315	leucokinin V - Mad
16	4	0.5	8	2 PT0725	T-cell receptor be
17	4	0.5	8	2 S68325	blood cell protein
18	3	0.4	3	3 PT0571	T-cell receptor be
19	3	0.4	4	2 A37832	phenol 2-monooxyge
20	3	0.4	4	2 S17255	ribosomal protein
21	3	0.4	4	2 JQ1273	neuropeptide Antho
22	3	0.4	4	2 E44823	synaptosomal-assoc
23	3	0.4	4	2 PT0633	T-cell receptor be
24	3	0.4	4	2 PT0711	T-cell receptor be
25	3	0.4	4	2 PT0661	T-cell receptor be
26	3	0.4	4	2 PT0698	T-cell receptor be
27	3	0.4	4	2 PT0677	T-cell receptor be
28	3	0.4	4	2 PT0706	T-cell receptor be
29	3	0.4	4	2 PT0675	T-cell receptor be

30	3	0.4	4	2 PT0721	T-cell receptor be
31	3	0.4	4	2 PT0566	T-cell receptor be
32	3	0.4	4	2 A26209	protein-glutamine
33	3	0.4	4	2 S55238	pallidipin - assas
34	3	0.4	5	2 C23751	spinal cord peptid
35	3	0.4	5	2 A41225	copper resistance
36	3	0.4	5	2 B41225	copper resistance
37	3	0.4	5	2 S70154	URE2 protein - Xan
38	3	0.4	5	2 I40698	biotin B - Citroba
39	3	0.4	5	2 A32014	tram protein - Esc
40	3	0.4	5	2 A44955	alkanol monooxygen
41	3	0.4	5	2 T14910	hypothetical prote
42	3	0.4	5	2 S11077	alpha-amylase - ri
43	3	0.4	5	2 PS0324	ribulose-bisphosph
44	3	0.4	5	2 B37988	acid proteinase li
45	3	0.4	5	2 B45525	actin I - malaria

ALIGNMENTS

RESULT 1

R-phycoerythrin beta-1 chain - red alga (Gastroclonium coulteri) (fragment)  
C22565  
C:Species: Gastroclonium coulteri  
C:Date: 07-Mar-1988 #sequence\_revision 07-Mar-1988 #text\_change 23-Mar-1993  
C:Accession: C22565  
R;Klotz, A.V.; Glazer, A.N.  
J. Biol. Chem. 260, 4856-4863, 1985  
A:Title: Characterization of the bilin attachment sites in R-phycoerythrin.  
A:Reference number: A22565; MUID:85182601; PMID:3886644  
A:Accession: C22565  
A:Molecule type: protein  
A:Residues: 1-6 <KLO>

Query Match 0.7%; Score 5; DB 2; Length 6;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 AACLR 121  
Db 2 AACLR 6

RESULT 2

PT0605  
T-cell receptor beta chain V-D-J region (120-11L) - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
C:Accession: PT0605  
R;Feeney, A.J.  
J. Exp. Med. 174, 115-124, 1991  
A:Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.  
A:Reference number: PT0509; MUID:91277601; PMID:1711558  
A:Accession: PT0605  
A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-6 <FEE>  
A:Experimental source: newborn thymus, strain BALB/c  
C:Keywords: T-cell receptor

Query Match 0.7%; Score 5; DB 2; Length 6;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 ASGAG 102  
Db 1 ASGAG 5

RESULT 3

PT0703  
T-cell receptor beta chain V-D-J region (135-11F) - mouse (fragment)

C;Species: Mus musculus (house mouse)  
 C;Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
 C;Accession: PT0703  
 R;Feeney, A.J.  
 J. Exp. Med. 174, 115-124, 1991  
 A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.  
 A;Reference number: PT0509; MUID:91277601; PMID:1711558  
 A;Accession: PT0703  
 A;Status: translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-5 <P>  
 A;Experimental source: newborn thymus, strain BALB/c  
 C;Keywords: T-cell receptor

Query Match 0.5%; Score 4; DB 2; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 496 SGD 499  
 Db 2 SGD 5

RESULT 4  
 PT0701  
 T-cell receptor beta chain V-D-J region - mouse (fragment)  
 C;Species: Mus musculus (house mouse)  
 C;Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
 C;Accession: PT0658; PT0701  
 R;Feeney, A.J.  
 J. Exp. Med. 174, 115-124, 1991  
 A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.  
 A;Reference number: PT0509; MUID:91277601; PMID:1711558  
 A;Accession: PT0658  
 A;Status: translation not shown  
 A;Molecule type: mRNA  
 A;Residues: 1-5 <P>  
 A;Experimental source: day 4 postnatal thymus, strain BALB/c, 121-3E  
 A;Accession: PT0701  
 A;Status: translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-5 <P>  
 A;Experimental source: newborn thymus, strain BALB/c, 161-2E  
 C;Keywords: T-cell receptor

Query Match 0.5%; Score 4; DB 2; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 652 SSGT 655  
 Db 2 SSGT 5

RESULT 5  
 A11490  
 pyruvate kinase (EC 2.7.1.40) - pig (fragment)  
 C;Species: Sus scrofa domestica (domestic pig)  
 C;Date: 05-Jun-1987 #sequence\_revision 05-Jun-1987 #text\_change 03-Mar-1995  
 C;Accession: A11490  
 R;Hjeltnist, G.; Andersson, J.; Edlund, B.; Engstrom, L.  
 Biochem. Biophys. Res. Commun. 61, 559-563, 1974  
 A;Title: Amino acid sequence of a (32-P)phosphopeptide from pig liver pyruvate kinase ph  
 A;Reference number: A11490; MUID:75127438; PMID:4375989  
 A;Accession: A11490  
 A;Molecule type: protein  
 A;Residues: 1-6 <P>  
 A;Experimental source: liver  
 C;Keywords: glycolysis; phosphotransferase

Query Match 0.5%; Score 4; DB 2; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 LRRA 117  
 Db 1 LRRA 4

## RESULT 6

B33932  
 Ig mu chain D region (D23) - mouse  
 C;Species: Mus musculus (house mouse)  
 C;Date: 09-Mar-1990 #sequence\_revision 09-Mar-1990 #text\_change 16-Aug-1996  
 C;Accession: B33932  
 R;Baccala, R.; Vo Quang, T.; Gilbert, M.; Ternynck, T.; Avrameas, S.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 4624-4628, 1989  
 A;Title: Two murine natural polyreactive autoantibodies are encoded by nonmutated germ-l  
 A;Reference number: A33932; MUID:89282823; PMID:2499887  
 A;Accession: B33932  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-6 <BAC>  
 A;Cross-references: GB:M27107  
 C;Keywords: immunoglobulin

Query Match 0.5%; Score 4; DB 2; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 LRLR 74  
 Db 3 LRLR 6

## RESULT 7

PT0604  
 T-cell receptor beta chain V-D-J region (111-1A) - mouse (fragment)  
 C;Species: Mus musculus (house mouse)  
 C;Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
 C;Accession: PT0604  
 R;Feeney, A.J.  
 J. Exp. Med. 174, 115-124, 1991  
 A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.  
 A;Reference number: PT0509; MUID:91277601; PMID:1711558  
 A;Accession: PT0604  
 A;Status: translation not shown  
 A;Molecule type: mRNA  
 A;Residues: 1-6 <P>  
 A;Experimental source: newborn thymus, strain BALB/c  
 C;Keywords: T-cell receptor

Query Match 0.5%; Score 4; DB 2; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 496 SGD 499  
 Db 2 SGD 5

## RESULT 8

PT0668  
 T-cell receptor beta chain V-D-J region (121-3BB) - mouse (fragment)  
 C;Species: Mus musculus (house mouse)  
 C;Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
 C;Accession: PT0668  
 R;Feeney, A.J.  
 J. Exp. Med. 174, 115-124, 1991  
 A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.  
 A;Reference number: PT0509; MUID:91277601; PMID:1711558  
 A;Accession: PT0668  
 A;Status: translation not shown  
 A;Molecule type: mRNA  
 A;Residues: 1-6 <P>  
 A;Experimental source: day 4 postnatal thymus, strain BALB/c

C;Keywords: T-cell receptor

Query Match 0.5%; Score 4; DB 2; Length 6;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 652 SSGT 655  
|||  
Db 2 SSGT 5

#### RESULT 9

JN0859  
peptidyl-di-peptidase A inhibitory peptide C105 - striped bonito  
C;Species: Sarda orientalis (striped bonito)  
C;Date: 10-Mar-1994 #sequence\_revision 10-Mar-1994 #text\_change 07-May-1999  
C;Accession: JN0859  
R;Matsumura, N.; Fujii, M.; Takeda, Y.; Shimizu, T.  
BioSci. Biotechnol. Biochem. 57, 1743-1744, 1993  
A;Title: Isolation and characterization of angiotensin I-converting enzyme inhibitory peptide  
A;Reference number: JN0859; PMID:94080036; PMID:7764272  
A;Accession: JN0859  
A;Molecule type: protein  
A;Residues: 1-7 <MA>  
A;Experimental source: intestine  
C;Comment: The carboxyl-terminus is essential for the protein's expression of angiotensin  
C;Superfamily: bradykinin-potentiating peptide  
C;Keywords: angiotensin-converting enzyme inhibitor

Query Match 0.5%; Score 4; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 158 KLEK 161  
|||  
Db 4 KLEK 7

#### RESULT 10

PT0654  
T-cell receptor beta chain V-D-J region (121-1BK) - mouse (fragment)  
C;Species: Mus musculus (house mouse)  
C;Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
C;Accession: PT0654  
R;Feeney, A.J.  
J. Exp. Med. 174, 115-124, 1991  
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.

A;Reference number: PT0509; PMID:91277601; PMID:1711558

A;Accession: PT0654  
A;Status: translation not shown  
A;Molecule type: mRNA  
A;Residues: 1-7 <FEE>  
A;Experimental source: day 4 postnatal thymus, strain BALB/c  
C;Keywords: T-cell receptor

Query Match 0.5%; Score 4; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 ASGA 101  
|||  
Db 1 ASGA 4

#### RESULT 11

PT0543  
T-cell receptor beta chain V-D-J region (126-1BE) - mouse (fragment)  
C;Species: Mus musculus (house mouse)  
C;Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997  
C;Accession: PT0543  
R;Feeney, A.J.  
J. Exp. Med. 174, 115-124, 1991  
A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.

A;Reference number: PT0509; PMID:91277601; PMID:1711558

A;Accession: PT0543

A;Status: translation not shown

A;Molecule type: mRNA

A;Residues: 1-7 <FEE>

A;Experimental source: day 18 fetal thymus, strain BALB/c

C;Keywords: T-cell receptor

Query Match 0.5%; Score 4; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 496 SGDQ 499  
|||  
Db 2 SGDQ 5

#### RESULT 12

PT0689

T-cell receptor beta chain V-D-J region (140-1AF) - mouse (fragment)

C;Species: Mus musculus (house mouse)

C;Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997

C;Accession: PT0689

R;Feeney, A.J.

J. Exp. Med. 174, 115-124, 1991

A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.

A;Reference number: PT0509; PMID:91277601; PMID:1711558

A;Accession: PT0689

A;Status: translation not shown

A;Molecule type: DNA

A;Residues: 1-7 <FEE>

A;Experimental source: day 18 fetal thymus, strain BALB/c

C;Keywords: T-cell receptor

Query Match 0.5%; Score 4; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 496 SGDQ 499  
|||  
Db 2 SGDQ 5

#### RESULT 13

PT0719

T-cell receptor beta chain V-D-J region (140-2F) - mouse (fragment)

C;Species: Mus musculus (house mouse)

C;Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 30-May-1997

C;Accession: PT0719; PT0638

R;Feeney, A.J.

J. Exp. Med. 174, 115-124, 1991

A;Title: Junctional sequences of fetal T cell receptor beta chains have few N regions.

A;Reference number: PT0509; PMID:91277601; PMID:1711558

A;Accession: PT0719

A;Status: translation not shown

A;Molecule type: DNA

A;Residues: 1-7 <FEE>

A;Experimental source: newborn thymus, strain BALB/c (clone 140-2F)

A;Accession: PT0638

A;Status: translation not shown

A;Molecule type: mRNA

A;Residues: 1-7 <FEE>

A;Experimental source: newborn thymus, strain BALB/c (clone 111-1N)

C;Keywords: T-cell receptor

Query Match 0.5%; Score 4; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 652 SSGT 655  
|||  
Db 2 SSGT 5

## RESULT 14

A32523  
 peptidyl-dipeptidase A (EC 3.4.15.1) - bovine (fragment)  
 N:Alternate names: angiotensin I-converting enzyme; peptidyl-dipeptidase I  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 18-Oct-1989 #sequence\_revision 18-Oct-1989 #text\_change 11-May-2000  
 C:Accession: A32523  
 R:Harris, R.B. Biol. 198, 513-521, 1986  
 Adv. Exp. Med. Biol. 198, 513-521, 1986  
 A:Title: Isolation and sequencing of an active-site peptide from angiotensin I-converting  
 A:Reference number: A32523; PMID:87123961; PMID:3028071  
 A:Accession: A32523  
 A:Molecule type: protein  
 A:Residues: 1-8 <HAR>  
 C:Superfamily: mammalian peptidyl-dipeptidase A  
 C:Keywords: alternative splicing; blood pressure control; peptidyl-dipeptide hydrolase; z

Query Match 0.5%; Score 4; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 TELA 287

Db 2 TELA 5

## RESULT 15

JS0315  
 leucokinin V - Madeira cockroach  
 C:Species: Leucophaea maderae (Madeira cockroach)  
 C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 20-Jun-2000  
 C:Accession: JS0315  
 R:Holman, G.M.; Cook, B.J.; Nachman, R.J.  
 Comp. Biochem. Physiol. C 88, 27-30, 1987  
 A:Title: Isolation, primary structure, and synthesis of leucokinins V and VI: myotropic  
 A:Reference number: JS0315  
 A:Accession: JS0315  
 A:Molecule type: protein  
 A:Residues: 1-8 <HOL>  
 C:Comment: Leucokinins, a family of cephalomyotropic peptides, stimulate contractile act  
 C:Keywords: amidated carboxyl end; cephalomyotropic peptide  
 F,8/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 0.5%; Score 4; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 650 GPSS 653

Db 3 GPSS 6

Search completed: August 17, 2004, 10:17:08  
 Job time : 21 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 17, 2004, 10:09:40 ; Search time 13 Seconds  
(without alignments)

2947.972 Million cell updates/sec

Title: US-10-045-815-4

Perfect score: 736

Sequence: 1 MAVRALKLLTLLAVAAAS.....PPEPAQESLSGSESKPKDEL 736

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 158

Minimum DB seq length: 0

Maximum DB seq length: 8

Post-processing: Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5	0.7	8	1 FAR4 MACRS	P83277 macrobrachi
2	4	0.5	5	1 TPIS CANFA	P54714 canis famil
3	4	0.5	8	1 LCK2 LEUNA	P21141 leucophaea
4	4	0.5	8	1 LCK5 LEUNA	P19987 leucophaea
5	3	0.4	4	1 ECSI HUMAN	P02731 homo sapien
6	3	0.4	4	1 FKA ANTEL	P58705 anthopleura
7	3	0.4	4	1 RM01 YEAST	P36515 saccharomyc
8	3	0.4	4	1 B105 CITR	P12997 citrobacter
9	3	0.4	5	1 PAP2 FARMA	P81864 pardachirus
10	3	0.4	5	1 PSK DAUCA	P58261 daucus caro
11	3	0.4	5	1 RE11 LITRU	P82070 litoria rub
12	3	0.4	5	1 RE21 LITRU	P82071 litoria rub
13	3	0.4	5	1 RE32 LITRU	P82073 litoria rub
14	3	0.4	5	1 SUGA ACHDO	P19991 acheta dome
15	3	0.4	5	1 TRM3 ECOLI	P13973 escherichia
16	3	0.4	5	1 UXA4 CHLTR	P38005 chlamydia t
17	3	0.4	6	1 ACPH RABIT	P25154 oryctolagus
18	3	0.4	6	1 ASP2 LACSN	P82655 lactobacill
19	3	0.4	6	1 FARP MONEX	P41966 moniezia ex
20	3	0.4	6	1 LOX1 LOCMT	P41491 locusta mig
21	3	0.4	6	1 UN06 CLOPA	P81351 clostridium
22	3	0.4	7	1 ALL2 CARMA	P81805 carcinus ma
23	3	0.4	7	1 BRHP CONIM	P58803 conus imper
24	3	0.4	7	1 CARP MYTEM	P10420 mytilus edu
25	3	0.4	7	1 CCF1 ENTEFA	P20104 enterococcu
26	3	0.4	7	1 CIA ENTEFA	P11932 enterococcu
27	3	0.4	7	1 GRFP MOUSE	P59025 mus musculu
28	3	0.4	7	1 LANC CARUI	P36960 carnobacter
29	3	0.4	7	1 MNP1 LEPEDE	P42984 leptinotars
30	3	0.4	7	1 PPH2 LYCES	P83379 lycopersisico
31	3	0.4	7	1 TPFY PACDA	P83455 pachymedusa
32	3	0.4	7	1 UF03 MOUSE	P38641 mus musculu
33	3	0.4	7	1 UH11 RAT	P56576 rattus norv

34	3	0.4	7	1 UN06 PINPS	P81675 pinus pinas
35	3	0.4	8	1 AKH MELML	P25423 melolontha
36	3	0.4	8	1 ALL5 CARMA	P81818 carcinus ma
37	3	0.4	8	1 ALL6 CARMA	P81819 carcinus ma
38	3	0.4	8	1 ALL1 CYDPO	P82152 cydia pomon
39	3	0.4	8	1 ALL5 CALVO	P41841 calliphora
40	3	0.4	8	1 ALL5 CYDPO	P82156 cydia pomon
41	3	0.4	8	1 ALL6 CYDPO	P82157 cydia pomon
42	3	0.4	8	1 ALL7 CARMA	P81809 carcinus ma
43	3	0.4	8	1 ALL8 CARMA	P81811 carcinus ma
44	3	0.4	8	1 ALL9 CARMA	P81812 carcinus ma
45	3	0.4	8	1 ANG2 BOTUA	Q10582 bothrops ja

#### ALIGNMENTS

RESULT 1  
FAR4 MACRS  
ID FAR4 MACRS STANDARD; PRT; 8 AA.  
AC P83277; 2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE FMRFamide-like neuropeptide FLP4 (APALRLRF-amide).  
OS Macrobrachium rosenbergii (Giant fresh water prawn).  
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Caridea;  
OC Palaemonoidea; Palaemonidae; Macrobrachium.  
OX NCBI\_TaxID=79674;  
RN [1]  
RC SEQUENCE, AND MASS SPECTROMETRY.  
RP TISSUE=Eyestalk;  
RA Sithigongkul P.; Saraithongkum W., Jaideechoey S., Longyant S.,  
RA Sithigongkul W.;  
RT "Novel FMRFamide-like neuropeptides from the eyestalk of the giant  
freshwater prawn Macrobrachium rosenbergii.";  
RL Comp. Biochem. Physiol. 120B:587-595(1998).  
CC -1- MASS SPECTROMETRY; MW=943; METHOD=MALDI.  
CC -1- SIMILARITY: Belongs to the FARP (FMRFamide related peptide)  
family.  
DR GO:0007218; P.neuropeptide signaling pathway; TAS.  
KW Neuropeptide; Amidation.  
FT MOD RES 8 8 AMIDATION.  
SQ SEQUENCE 8 AA; 943 MW; 9CD40734072DC76D CRC64;

Query Match 0.7%; Score 5; DB 1; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 70 ALRLR 74  
| | | | |  
Db 3 ALRLR 7

RESULT 2  
TPIS CANFA  
ID TPIS CANFA STANDARD; PRT; 5 AA.  
AC P54714;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Triosephosphate isomerase (EC 5.3.1.1) (TIM) (Fragment).  
GN TP11.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RC SEQUENCE.  
RP TISSUE=Heart;  
RC MEDLINE=98163340; PubMed=9504812;  
RA Dunn M.J., Corbett J.M., Wheeler C.H.;

RT "HSC-2DPAGE and the two-dimensional gel electrophoresis database of  
 RT dog heart proteins.";  
 RL Electrophoresis 18:2795-2802(1997).  
 CC -!- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate = glyceralone  
 CC phosphate.  
 CC -!- PATHWAY: Plays an important role in several metabolic pathways.  
 CC -!- SUBUNIT: Homodimer (By similarity).  
 CC -!- SIMILARITY: Belongs to the triosephosphate isomerase family.  
 DR HSC-2DPAGE; P54714; DOG.  
 DR InterPro; IPR000652; Triophos.ismrse.  
 DR PROSITE; PS00171; TIM; PARTIAL.  
 KW Isomerase; Glycolysis; Gluconeogenesis; Fatty acid biosynthesis;  
 KW Pentose shunt.  
 FT NON\_TER 1 1  
 FT NON\_TER 5 5  
 SQ SEQUENCE 5 AA; 550 MW; 64444862C9A00000 CRC64;

Query Match 0.5%; Score 4; DB 1; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 170 FVGN 173  
 Db ||||  
 1 FVGN 4

RESULT 3  
 LCK2\_LEUMA  
 ID LCK2\_LEUMA STANDARD; PRT; 8 AA.  
 AC P21141;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Leucokinin II (L-II);  
 OS Leucophaea maderae (Madeira cockroach).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Orthopteroidea; Dictyoptera; Blattaria; Blaberoidea;  
 OC Blaberoidea; Leucophaea.  
 OX NCBI\_TaxID=6988;  
 RN [1]  
 RP SEQUENCE, AND SYNTHESIS.  
 RC TISSUE=Head;  
 RA Holman G.M., Cook B.J., Nachman R.J.;  
 RT "Isolation, primary structure and synthesis of two neuropeptides  
 RT from Leucophaea maderae: members of a new family of  
 RT Cephalomyotropins.";  
 RL Comp. Biochem. Physiol. 84C:205-211(1986).  
 CC -!- FUNCTION: This cephalomyotropic peptide stimulates contractile  
 CC activity of cockroach proctodeum (hindgut).  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 KW Neuropeptide; Amidation.  
 FT MOD\_RES 8 8  
 FT MOD\_RES 8 8  
 SQ SEQUENCE 8 AA; 852 MW; DC6365A5B9C8676A CRC64;

Query Match 0.5%; Score 4; DB 1; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 650 GFSS 653  
 Db ||||  
 3 GFSS 6

RESULT 4  
 LCK5\_LEUMA  
 ID LCK5\_LEUMA STANDARD; PRT; 8 AA.  
 AC P19987;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Leucokinin V (L-V).  
 OS Leucophaea maderae (Madeira cockroach).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Orthopteroidea; Dictyoptera; Blattaria; Blaberoidea;  
 OC Blaberoidea; Leucophaea.  
 OX NCBI\_TaxID=6988;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Head;  
 RA MEDLINE=87052651; PubMed=2877794;  
 RA Holman G.M., Cook B.J., Nachman R.J.;  
 RT "Isolation, primary structure, and synthesis of leucokinin V and VI:  
 RT myotropic peptides of Leucophaea maderae.";  
 RL Comp. Biochem. Physiol. 88C:27-30(1987).  
 CC -!- FUNCTION: This cephalomyotropic peptide stimulates contractile  
 CC activity of cockroach proctodeum (hindgut).  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 DR PIR; JS0315; JS0315.  
 DR Neuropeptide; Amidation.  
 KW MOD\_RES 8 8  
 FT MOD\_RES 8 8  
 SQ SEQUENCE 8 AA; 784 MW; 736365A5B9C865B8 CRC64;

Query Match 0.5%; Score 4; DB 1; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 650 GFSS 653  
 Db ||||  
 3 GFSS 6

RESULT 5  
 EOSI\_HUMAN  
 ID EOSI\_HUMAN STANDARD; PRT; 4 AA.  
 AC P02731;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 21-JUL-1986 (Rel. 01, Last annotation update)  
 DE Eosinophilotactic peptides.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE.  
 RC MEDLINE=76078412; PubMed=1060093;  
 RA Goetzi E.J., Austen K.F.;  
 RT "Purification and synthesis of eosinophilotactic tetrapeptides of  
 RT human lung tissue: identification as eosinophil chemotactic factor of  
 RT anaphylaxis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 72:4123-4127(1975).  
 CC -!- MISCELLANEOUS: These peptides are released from mast cells in lung  
 CC (and other tissues) during hypersensitivity reactions  
 CC (anaphylaxis); their activities, preferentially affecting  
 CC eosinophils, include chemotaxis, chemotactic deactivation, release  
 CC of enzymes, and stimulation of the hexose monophosphate shunt.  
 DR GO; GO:0006935; P:chemotaxis; IDA.  
 DR GO; GO:0006955; P:immune response; IDA.  
 DR GO; GO:0006955; P:immune response; IDA.  
 FT VARIANT 1 1  
 FT VARIANT V -> A (IN OTHER PEPTIDE).  
 FT FTID=VAR 005201.  
 SQ SEQUENCE 4 AA; 390 MW; 6B05B862A0000000 CRC64;

Query Match 0.4%; Score 3; DB 1; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 727 GSE 729  
 Db ||||  
 2 GSE 4

RESULT 6  
 PFKA\_ANTEL  
 ID PFKA\_ANTEL STANDARD; PRT; 4 AA.  
 AC P58705;  
 DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DE Antho-KAamide.  
 OS Anthopleura elegantissima (Sea anemone).  
 OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;  
 OC Nynanthaeae; Actiniidae; Anthopleura.  
 CX NCBI\_TaxID=6110;  
 RN [1]  
 RP SEQUENCE  
 RX MEDLINE=92028852; PubMed=1681803;  
 RA Nothacker H.-P., Rinehart K.L. Jr., Grimmelikhuijzen C.J.P.;  
 RT "Isolation of L-3-phenylactyl-Phe-Lys-Ala-NH2 (Antho-KAamide), a  
 RT novel neuropeptide from sea anemones."  
 RL Biochem. Biophys. Res. Commun. 179:1205-1211(1991).  
 RN [2]  
 RP FUNCTION  
 RX MEDLINE=93391436; PubMed=8397415;  
 RA McFarlane I.D., Hudman D., Nothacker H.-P., Grimmelikhuijzen C.J.P.;  
 RT "The expansion behaviour of sea anemones may be coordinated by two  
 RT inhibitory neuropeptides, Antho-KAamide and Antho-Riamide."  
 RL Proc. R. Soc. Lond., B, Biol. Sci. 253:183-188(1993).  
 CC -!- FUNCTION: Inhibits spontaneous contractions in several muscle  
 CC groups. May be involved in the expansion phase of feeding  
 CC behaviour in sea anemones.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Neuron specific.  
 DR PIR; JQ1273; JQ1273.  
 KW Neuropeptide; Amidation.  
 FT MOD RES 1 1 L-3-PHENYLACTYL.  
 FT MOD RES 4 4 AMIDATION.  
 SQ SEQUENCE 4 AA; 512 MW; 6DD339CA00000000 CRC64;  
  
 Query Match 0.4%; Score 3; DB 1; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 519 FKA 521  
 DB 2 FKA 4  
  
 RESULT 7  
 RM01\_YEAST STANDARD; PRT; 4 AA.  
 AC P36515;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DE 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 01-JUN-1994 (Rel. 29, Last annotation update)  
 DE Mitochondrial 60S ribosomal protein L1 (YmL1) (Fragment).  
 GN MRPL1.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 CX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=91285106; PubMed=2060626;  
 RA Grobmann L., Graack H.-R., Kruff V., Choli T., Goldschmidt-Reisin S.,  
 RA Kitakawa M.;  
 RT "Extended N-terminal sequencing of proteins of the large ribosomal  
 RT subunit from yeast mitochondria."  
 RL FEBS Lett. 284:51-56(1991).  
 DR PIR; S17255; S17255.  
 DR SGD; L0002681; MRPL1.  
 KW Ribosomal protein; Mitochondrion.  
 FT NON\_TER 4 4  
 SQ SEQUENCE 4 AA; 402 MW; 7771B2D5D000000000 CRC64;  
  
 Query Match 0.4%; Score 3; DB 1; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 33 VTP 35

Db 2 VTP 4  
  
 RESULT 8  
 BIOB\_CITFR STANDARD; PRT; 5 AA.  
 ID BIOB\_CITFR  
 AC P12997;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Biotin synthase (EC 2.8.1.6) (Biotin synthetase) (Fragment).  
 GN BIOB.  
 OS Citrobacter freundii.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Citrobacter.  
 CX NCBI\_TaxID=546;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89006280; PubMed=2371595;  
 RA Shiuan D., Campbell A.;  
 RT "Transcriptional regulation and gene arrangement of Escherichia coli,  
 RT Citrobacter freundii and Salmonella typhimurium biotin operons."  
 RL Gene 67:203-211(1988).  
 CC -!- CATALYTIC ACTIVITY: Dethiobiotin + sulfur = biotin.  
 CC -!- PATHWAY: Biotin biosynthesis; last step.  
 CC -!- SIMILARITY: Belongs to the biotin and lipoic acid synthetases  
 CC family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; M21922; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; I40698; I40698.  
 KW Biotin biosynthesis; Iron-sulfur; Transferase.  
 FT NON\_TER 5 5  
 SQ SEQUENCE 5 AA; 532 MW; 75A5B1EDD6F00000 CRC64;  
  
 Query Match 0.4%; Score 3; DB 1; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 129 AHS 131  
 DB 2 AHS 4  
  
 RESULT 9  
 PAP2\_PARMA STANDARD; PRT; 5 AA.  
 ID PAP2\_PARMA  
 AC P81864;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Pardaxin II (PXII) (Fragment).  
 OS Pardachirus marmoratus (Red sea moose sole).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;  
 OC Soleiidae; Soleidae; Pardachirus.  
 CX NCBI\_TaxID=31087;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE: Skin secretion;  
 RX MEDLINE=87057369; PubMed=3782138;  
 RA Lazarovici P., Primor N., Loew L.M.;  
 RT "Purification and pore-forming activity of two hydrophobic  
 RT polypeptides from the secretion of the Red sea moose sole (Pardachirus

```

RT marmoratus).";
RL J. Biol. Chem. 261:16704-16713(1986).
CC -!- FUNCTION: Exhibits unusual shark repellent and surfactant
CC properties. Forms voltage-dependent, ion-permeable channels
CC in membranes. At high concentration causes cell membrane lysis.
CC -!- SUBUNIT: Monomer. In aqueous solution exists as a tetramer.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the pardaxin family.
KW Toxin.
FT NON_TER
SQ SEQUENCE 5 AA; 614 MW; 7769C9C9C8100000 CRC64;

Query Match 0.4%; Score 3; DB 1; Length 5;
Best Local Similarity 100.0%; Pred.No.1.4e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 PFP 332
Db |||
3 PFP 5

RESULT 10
PSK_DAUCA
ID PSK_DAUCA STANDARD; PRT; 5 AA.
AC P58261.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phytosulfokine-alpha (PSK-alpha) [Contains: Phytosulfokine-beta (PSK-beta)].
OS Daucus carota (Carrot).
CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
CC campanulids; Apiales; Apiaceae; Apioidae; Scandiceae; Daucinae;
CC Daucus.
OX NCBI_TaxID=4039;
RN [1]
RP SEQUENCE, AND IDENTIFICATION BY MASS SPECTROMETRY.
RC STRAIN=cv. US-Harumakigosun;
RX MEDLINE=20212743; PubMed=10750705;
RA Hanai H., Matsuno T., Yamamoto M., Matsubayashi Y., Kobayashi T.,
RA Kamada H., Sakagami Y.
RT "A secreted peptide growth factor, phytosulfokine, acting as a
RT stimulatory factor of carrot somatic embryo formation.";
RL Plant Cell Physiol. 41:27-32(2000).
CC -!- FUNCTION: In presence of 2,4-D, stimulates proliferation of the
CC cells, but does not stimulate differentiation into the somatic
CC embryos.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- PTM: Sulfation is important for activity and for the binding to a
CC putative membrane receptor (By similarity).
CC -!- SIMILARITY: Belongs to the phytosulfokine family.
KW Growth factor; Sulfation.
FT PEPTIDE 1 4 PHYTOSULFOKINE-BETA.
FT MOD RES 1 1 SULFATION.
FT MOD RES 3 3 SULFATION.
SQ SEQUENCE 5 AA; 687 MW; 76C1BB504B300000 CRC64;

Query Match 0.4%; Score 3; DB 1; Length 5;
Best Local Similarity 100.0%; Pred.No.1.4e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 317 YQ 319
Db |||
3 YQ 5

RESULT 11
RE11_LITRU
ID RE11_LITRU STANDARD; PRT; 5 AA.
AC P82070;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)

```

```

DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Rubellidin 1.1.
OS Litoria rubella (Desert tree frog).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Amphibia; Batrachia; Anura; Neobatrachia; Hylidae; Hylinae;
CC Pelodyadinae; Litoria.
OX NCBI_TaxID=104895;
RN [1]
RP SEQUENCE, AND MASS SPECTROMETRY.
RC TISSUE=Skin secretion;
RA Steinbörner S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,
RA Tyler M.J., Wallace J.C.;
RT "The structure of new peptides from the Australian red tree frog
RT 'Litoria rubella'. The skin peptide profile as a probe for the study
RT of evolutionary trends of amphibians.";
RL Aust. J. Chem. 49:955-963(1996).
CC -!- FUNCTION: Shows neither neuropeptide activity nor antibiotic
CC activity.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Expressed by the skin dorsal glands.
CC -!- MASS SPECTROMETRY: MW=598; METHOD=FAB.
KW Amphibian defense peptide.
SQ SEQUENCE 5 AA; 598 MW; 6DD9C9CAB2A00000 CRC64;

Query Match 0.4%; Score 3; DB 1; Length 5;
Best Local Similarity 100.0%; Pred.No.1.4e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 377 PFA 379
Db |||
3 PFA 5

RESULT 12
RE21_LITRU
ID RE21_LITRU STANDARD; PRT; 5 AA.
AC P82071;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Rubellidin 2.1.
OS Litoria rubella (Desert tree frog).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Amphibia; Batrachia; Anura; Neobatrachia; Hylidae; Hylinae;
CC Pelodyadinae; Litoria.
OX NCBI_TaxID=104895;
RN [1]
RP SEQUENCE, AND MASS SPECTROMETRY.
RC TISSUE=Skin secretion;
RA Steinbörner S.T., Wabnitz P.A., Waugh R.J., Bowie J.H., Gao C.,
RA Tyler M.J., Wallace J.C.;
RT "The structure of new peptides from the Australian red tree frog
RT 'Litoria rubella'. The skin peptide profile as a probe for the study
RT of evolutionary trends of amphibians.";
RL Aust. J. Chem. 49:955-963(1996).
CC -!- FUNCTION: Shows neither neuropeptide activity nor antibiotic
CC activity.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Expressed by the skin dorsal glands.
CC -!- MASS SPECTROMETRY: MW=626; METHOD=FAB.
KW Amphibian defense peptide.
SQ SEQUENCE 5 AA; 626 MW; 6DD9C9C9CB10300000 CRC64;

Query Match 0.4%; Score 3; DB 1; Length 5;
Best Local Similarity 100.0%; Pred.No.1.4e+05;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 377 PFA 379
Db |||
3 PFA 5

RESULT 13

```



RE32\_LITRU STANDARD; PRT; 5 AA.  
ID RE32\_LITRU  
AC P82073;  
DT 28-FEB-2003 (Rel. 41, Created)  
DE TRAM protein (Fragment).  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE TRAM  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Rubellidin 3.2.  
OS Litoria rubella (Desert tree frog).  
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Amphibia; Batrachia; Anura; Neobatrachia; Hylidae; Hylinae;  
CC Pelodyadinae; Litoria.  
OX NCBI\_TaxID=104895;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88227859; PubMed=2836369;  
RA Inamoto S., Yoshioka Y., Ohtsubo E.;  
RT "Identification and characterization of the products from the traJ  
and traY genes of plasmid Ri00.";  
RL J. Bacteriol. 170:2749-2757(1988).  
CC -!- FUNCTION: TRANSFER GENE PROTEIN. IS INVOLVED IN THE CONJUGATION  
PROCESS OF BACTERIAL CELLS FOR THE EXCHANGE OF PLASMID DNA.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: Belongs to the traM family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M20941; -; NOT\_ANNOTATED\_CDS.  
DR PIR; A32014; A32014.  
KW Conjugation; Plasmid; DNA-binding.  
FT NON\_TER 1  
SQ SEQUENCE 5 AA; 570 MW; 71A9C9C862A00000 CRC64;  
  
Query Match 0.4%; Score 3; DB 1; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 649 VGF 651  
DB 1 VGF 3  
  
RESULT 14  
SUGA\_ACHDO STANDARD; PRT; 5 AA.  
ID SUGA\_ACHDO  
AC F19991;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Subesophageal ganglion pentapeptide.  
OS Acheta domesticus (House cricket).  
CC Eukaryota; Metazoa; Arthropoda; Insecta; Pserygota;  
CC Neoptera; Orthopteroidea; Orthoptera; Ensifera; Gryllidae; Gryllinae;  
CC Acheta.  
OX NCBI\_TaxID=6937;  
RN [1]  
RP SEQUENCE.  
RA Wicker C., Wicker C.;  
RT "Isolation and structure of a peptide isolated from the  
subesophageal ganglion of Acheta domesticus (Orthoptera).";  
RL Comp. Biochem. Physiol. 88C:185-187(1987).  
CC -!- SUBCELLULAR LOCATION: Main peptide from the subesophageal  
ganglia.  
CC  
DR PIR; JS0319; JS0319.  
SQ SEQUENCE 5 AA; 476 MW; 69D76DDDDDB00000 CRC64;  
  
Query Match 0.4%; Score 3; DB 1; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 17 AAA 19  
DB 1 AAA 3  
  
RESULT 15  
TRM3\_ECOLI STANDARD; PRT; 5 AA.  
ID TRM3\_ECOLI  
AC F13973;

DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE TRAM protein (Fragment).  
GN TRAM  
OS Escherichia coli.  
OG Plasmid IncFII Ri00.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88227859; PubMed=2836369;  
RA Inamoto S., Yoshioka Y., Ohtsubo E.;  
RT "Identification and characterization of the products from the traJ  
and traY genes of plasmid Ri00.";  
RL J. Bacteriol. 170:2749-2757(1988).  
CC -!- FUNCTION: TRANSFER GENE PROTEIN. IS INVOLVED IN THE CONJUGATION  
PROCESS OF BACTERIAL CELLS FOR THE EXCHANGE OF PLASMID DNA.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: Belongs to the traM family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M20941; -; NOT\_ANNOTATED\_CDS.  
DR PIR; A32014; A32014.  
KW Conjugation; Plasmid; DNA-binding.  
FT NON\_TER 1  
SQ SEQUENCE 5 AA; 634 MW; 6B1B1AA443500000 CRC64;  
  
Query Match 0.4%; Score 3; DB 1; Length 5;  
Best Local Similarity 100.0%; Pred. No. 1.4e+05;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 333 NDE 335  
DB 2 NDE 4  
  
Search completed: August 17, 2004, 10:15:44  
Job time : 14 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 17, 2004, 10:12:45 ; Search time 42 Seconds  
(without alignments)  
5529,081 Million cell updates/sec

Title: US-10-045-815-4  
Perfect score: 736  
Sequence: 1 MAVRALKLTTLAVVAAS.....PPEPAQESLSGSEKPKDEL 736

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 488

Minimum DB seq length: 0

Maximum DB seq length: 8

Post-processing: Listing first 45 summaries

Database : SPTREMBL 25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mbc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4	0.5	7	Q54248	Q54248 streptomyce
2	4	0.5	8	O87471	O87471 haemophilus
3	4	0.5	8	Q51594	Q51594 escherichia
4	4	0.5	8	Q72719	Q72719 homo sapien
5	4	0.5	8	Q94623	Q94623 manduca sex
6	4	0.5	8	Q40659	Q40659 oryza sativ
7	4	0.5	8	Q42507	Q42507 triticum ae
8	4	0.5	8	P82324	P82324 pisum sativ
9	3	0.4	6	P83569	P83569 sepia offic
10	3	0.4	6	P82541	P82541 spinacia ol
11	3	0.4	7	Q8KMS3	Q8KMS3 klebsiella
12	3	0.4	7	Q47477	Q47477 escherichia
13	3	0.4	7	P72081	P72081 nocardia la
14	3	0.4	7	Q8KMS9	Q8KMS9 enterobacte
15	3	0.4	7	Q07354	Q07354 synechococc
16	3	0.4	7	P83530	P83530 lactobacill

17	3	0.4	7	4	Q8NH7	Q8nh7 homo sapien
18	3	0.4	7	5	Q7Z1C0	Q7z1c0 caenorhabdi
19	3	0.4	7	8	O9182	O9182 gnatholebia
20	3	0.4	7	8	Q95945	Q95945 saccharomyc
21	3	0.4	7	10	O49223	O49223 glycine max
22	3	0.4	7	10	P82445	P82445 nicotiana t
23	3	0.4	7	11	O8K3H6	O8k3h6 rattus norv
24	3	0.4	7	11	O63668	O63668 rattus norv
25	3	0.4	7	12	Q9YVE3	Q9yve3 human adeno
26	3	0.4	7	12	Q66205	Q66205 transmissib
27	3	0.4	7	12	Q66113	Q66113 cherry leaf
28	3	0.4	7	12	Q9YIQ9	Q9yiq9 human adeno
29	3	0.4	7	12	Q9YQ10	Q9yq10 transmissib
30	3	0.4	7	12	Q9YIR0	Q9yir0 human adeno
31	3	0.4	7	13	O8J120	O8j120 gallus gall
32	3	0.4	7	13	O42564	O42564 fugu rubrip
33	3	0.4	8	2	Q9S6D5	Q9s6d5 escherichia
34	3	0.4	8	2	P72279	P72279 rhodococcus
35	3	0.4	8	2	Q8KPX4	Q8kpx4 microcystis
36	3	0.4	8	2	O68485	O68485 klebsiella
37	3	0.4	8	2	Q93SR0	Q93sr0 staphylococ
38	3	0.4	8	2	O87J10	O87j10 streptomyce
39	3	0.4	8	2	Q92EZ9	Q92ez9 buchnera ap
40	3	0.4	8	2	Q56759	Q56759 xanthobacte
41	3	0.4	8	2	O32560	O32560 escherichia
42	3	0.4	8	2	O09258	O09258 synechococc
43	3	0.4	8	2	Q9X3K1	Q9x3k1 prochloroco
44	3	0.4	8	2	Q9ZIE9	Q9zie9 neisseria m
45	3	0.4	8	2	Q9R9E0	Q9r9e0 bacillus su

#### ALIGNMENTS

##### RESULT 1

Q54248 ID Q54248 PRELIMINARY; PRT; 7 AA.  
AC Q54248;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE RplO protein (Fragment).  
GN RPL0.  
OS Streptomyces griseus.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OC NCBI\_TaxID=1911;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N2-3-11;  
RX MEDLINE=20011291; PubMed=10542330;  
RA Poehling S., Piepersberg W., Wehmeier U.F.;  
RT "Analysis and regulation of the sec Y gene from Streptomyces griseus  
N2-3-11 and interaction of the SecY protein with the SecA protein.";  
RL Biochim. Biophys. Acta 1447:298-302(1999).  
DR EMBL; X95915; CAA65160.1; -;  
FT NON\_TER  
SQ SEQUENCE 7 AA; 760 MW; 72C72B01B2D1B2A0 CRC64;

Query Match 0.5%; Score 4; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1e+06;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 283 VTEL 286  
|||  
Db 3 VTEL 6

##### RESULT 2

O87471 ID O87471 PRELIMINARY; PRT; 8 AA.  
AC O87471;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)

DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update).  
 DE HIFA (Fragment).  
 GN HIFA.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]\_TaxID=727;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Eagan;  
 RX MEDLINE=98389689; PubMed=97211313;  
 RA Mhlanga-Mutagadara T., Morlin G., Smith A.L., Eisenstark A.,  
 RA Golomb M.;  
 RA "Evolution of the major pilus gene cluster of haemophilus  
 RT influenzae.";  
 RL J. Bacteriol. 180:4693-4703(1998).  
 DR EMBL; AF071762; AAC35830.1; -.  
 FT NON TER 1  
 SQ SEQUENCE 8 AA; 876 MW; DAB44451A7272325 CRC64;

Query Match 0.5%; Score 4; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 KLLT 10  
 DB 2 KLLT 5

## RESULT 3

Q51594 PRELIMINARY; PRT; 8 AA.  
 AC Q51594;  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
 DE CopB protein (Fragment).  
 OS Escherichia coli.  
 OG Plasmid ColV2-K94.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86223772; PubMed=2423502;  
 RA Weber P.C., Palchaudhuri S.;  
 RT "Incompatibility repressor in a repA-like replicon of the IncFI  
 RT plasmid ColV2-K94.";  
 RL J. Bacteriol. 166:1106-1112(1986).  
 DR EMBL; M13472; AAA23194.1; -.  
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
 KW Plasmid.  
 FT NON TER 1  
 SQ SEQUENCE 8 AA; 998 MW; 0F37205AA73416D7 CRC64;

Query Match 0.5%; Score 4; DB 2; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 486 LQRL 489  
 DB 1 LQRL 4

## RESULT 4

Q72719 PRELIMINARY; PRT; 8 AA.  
 AC Q72719;  
 DT 01-OCT-2003 (TReMBLrel. 25, Created)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
 DE Retinoblastoma 1 (Fragment).

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Leone P.E., Paz-y-Mino C.S., Jervis P., Vega M.A.E., Pestana A.,  
 RA Alonso J.;  
 RA "Polymorphism of the RB1 gene in intron 16";  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY260473; AAP39868.1; -.  
 FT NON TER 1  
 FT NON TER 8  
 SQ SEQUENCE 8 AA; 821 MW; DC65BAA72456C5A1 CRC64;

Query Match 0.5%; Score 4; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 QNLD 184  
 DB 3 QNLD 6

## RESULT 5

Q94623 PRELIMINARY; PRT; 8 AA.  
 AC Q94623;  
 DT 01-FEB-1997 (TReMBLrel. 02, Created)  
 DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)  
 DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)  
 DE MsUSP-2 protein (Fragment).  
 GN USP.  
 OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Spingioidea;  
 OC Spingidae; Spinginae; Manduca.  
 OX NCBI\_TaxID=71130;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Epidermis;  
 RX MEDLINE=97165493; PubMed=9013254;  
 RA Jindra M., Huang J.Y., Malone F., Asahina M., Riddiford L.M.;  
 RT "Identification and mRNA developmental profiles of two ultraspiracle  
 RT isoforms in the epidermis and wings of Manduca sexta.";  
 RL Insect Mol. Biol. 6:41-53(1997).  
 DR EMBL; U57921; AAB64235.1; -.  
 FT NON TER 8  
 SQ SEQUENCE 8 AA; 892 MW; F165BB0415A76B16 CRC64;

Query Match 0.5%; Score 4; DB 5; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1e+06;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 290 PSRE 293  
 DB 3 PSRE 6

## RESULT 6

Q40659 PRELIMINARY; PRT; 8 AA.  
 AC Q40659;  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)  
 DE Alpha-amylase (Fragment).  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]

```

RP SEQUENCE FROM N.A.
RX MEDLINE=91078641; PubMed=2258052;
RA Kumagai M.H., Shah M., Terashima M., Vrkljan Z., Whitaker J.R.,
RA Rodriguez R.L.;
RT "Expression and secretion of rice alpha-amylase by saccharomycers
RT cerevisiae.";
RL Gene 94:209-216(1990).
DR EMBL; M62916; AAA33892.1; -.
DR Gramene; Q40659; -.
FT NON TER 8
SQ SEQUENCE 8 AA; 948 MW; EBC694444732D6D6 CRC64;

Query Match 0.5%; Score 4; DB 10; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 274 QVLN 277
Db 2 QVLN 5

RESULT 7
Q42507 PRELIMINARY; PRT; 8 AA.
ID AC Q42507;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Heat shock protein (Fragment).
GN HSP70C OR HSP70A OR HSP70B.
OS Triticum aestivum (wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Triticum.
OC NCBI_TaxID=4565;

[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=96189275; PubMed=8605312;
RX STRAIN=CV MUSTANG;
RA Joshi C.P., Kumar S., Nguyen H.T.;
RT "Application of modified differential display technique for cloning
RT and sequencing of the 3' region from three putative members of wheat
RT HSP70 gene family.";
RL Plant Mol. Biol. 30:641-646(1996).
DR EMBL; L41507; AAB02333.1; -.
DR EMBL; L41505; AAB02331.1; -.
DR EMBL; L41506; AAB02332.1; -.
DR CO; GO:0003773; F:heat shock protein activity; IEA.
KW Heat shock.
FT NON TER 1
SQ SEQUENCE 8 AA; 886 MW; 71B2CB1B10532768 CRC64;

Query Match 0.5%; Score 4; DB 10; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 572 IEVV 575
Db 4 IEVV 7

RESULT 8
P82324 PRELIMINARY; PRT; 8 AA.
ID AC P82324;
DT 01-JUN-2000 (TrEMBLrel. 14, Created)
DT 01-JUN-2000 (TrEMBLrel. 14, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Unknown protein from 2D-page of thylakoid (SPOT105) (Fragment).
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.

```

```

OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE, SUBCELLULAR LOCATION, AND DEVELOPMENTAL STAGE.
RC STRAIN=CV. DE GRACE; TISSUE=LEAF;
RX MEDLINE=20181728; PubMed=10715320;
RA Peltier J.-B., Friso G., Kalume D.E., Roepstorff P., Nilsson F.,
RA Adamska I., van Wijk K.J.;
RT "Proteomics of the chloroplast: systematic identification and
RT targeting analysis of luminal and peripheral thylakoid proteins.";
RL Plant Cell 12:319-341(2000).
CC -|- SUBCELLULAR LOCATION: CHLOROPLAST THYLAKOID MEMBRANE LUMEN OR
CC PERIPHERY.
CC -|- DEVELOPMENTAL STAGE: UNFOLDED AND FULLY DEVELOPED LEAVES.
CC -|- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC PROTEIN IS: 6.7, ITS MW IS: 16.8 KDA.
DR GO; GO:0009507; C:chloroplast; IEA.
DR GO; GO:0009579; C:thylakoid; IEA.
KW Chloroplast; Thylakoid.
FT NON TER 1
FT NON TER 8
SQ SEQUENCE 8 AA; 839 MW; DDC68B5DDDC2D2D5 CRC64;

Query Match 0.5%; Score 4; DB 10; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 VVAA 18
Db 2 VVAA 5

RESULT 9
P83569 PRELIMINARY; PRT; 6 AA.
ID AC P83569;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Sperm attracting peptide SepSAF.
OS Sepia officinalis (Common cuttlefish).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;
OC Decapodiformes; Sepioidea; Sepiidae; Sepia.
OC NCBI_TaxID=6610;

[1]
RN SEQUENCE, FUNCTION, DEVELOPMENTAL STAGE, MASS SPECTROMETRY, AND
RP AMIDATION.
RC TISSUE=Egg;
RX Zatylny C., Marvin L., Gagnon J., Henry J.;
RT "Fertilization in Sepia officinalis: the first mollusk sperm-
RT attracting peptide.";
RL Biochem. Biophys. Res. Commun. 296:1186-1193(2002).
CC -|- FUNCTION: ATTRACTS SPERM INCREASING THE CHANCES OF GAMETE
CC COLLISION.
CC -|- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE OVARIAN FOLLICLES FURING
CC VITELLOGENESIS. ACCUMULATES IN THE OOCYTES BEFORE BEING SECRETED
CC DURING FERTILIZATION. EXPRESSION CONTINUES IN THE EMBEDDED OOCYTE.
CC ACCUMULATES IN THE EGG CAPSULE AFTER FERTILIZATION.
CC -|- MASS SPECTROMETRY: MW=556.6; METHOD=MALDI.
KW Amidation.
FT MOD RES 6
FT MOD RES 6
SQ SEQUENCE 6 AA; 597 MW; 72C8676AA0470000 CRC64;

Query Match 0.4%; Score 3; DB 5; Length 6;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 PGV 54
Db 4 PGV 6

RESULT 10

```

P82541  
 ID P82541 PRELIMINARY; PRT; 6 AA.  
 AC P82541;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Chloroplast 30S ribosomal protein S19 beta (Fragment).  
 OS Spinacia oleracea (Spinach).  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Caryophyllales; Amaranthaceae; Spinacia.  
 OX NCBI\_TaxID=3362;  
 [1]  
 RP SEQUENCE, FUNCTION, AND MASS SPECTROMETRY.  
 RC STRAIN=cv. ALVARO; TISSUE=Leaf;  
 RX MEDLINE=20435797; PubMed=10874039;  
 RA Yanaguchi K., von Knoblauch K., Subramanian A.R.;  
 RT "The plastid ribosomal proteins. Identification of all the proteins in  
 RT the small subunit of an organelle ribosome (Chloroplast).";  
 RL J. Biol. Chem. 37:28455-28465(2000).  
 CC -!- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 16S RIBOSOMAL RNA.  
 CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN ALL PLANT TISSUES.  
 CC -!- MASS SPECTROMETRY: MW=10477.0; METHOD=ELECTROSPRAY.  
 CC -!- MASS SPECTROMETRY: MW=10495; METHOD=MALDI  
 CC -!- MISCELLANEOUS: S19 ALPHA AND BETA FORMS DIFFER IN PI. S19 BETA  
 CC FORM IS THE MINOR BASIC FORM.  
 CC -!- MISCELLANEOUS: ON THE 2D-GEL ITS MW IS: 12 KDA.  
 CC -!- SIMILARITY: BELONGS TO THE S19P FAMILY OF RIBOSOMAL PROTEINS.  
 DR GO; GO:0009507; C:chloroplast; IEA.  
 DR GO; GO:0019843; F:rRNA binding; IEA.  
 DR GO; GO:0003735; P:structural constituent of ribosome; IEA.  
 DR InterPro; IPR002222; Ribosomal\_S19.  
 DR PROSITE; PS00323; RIBOSOMAL\_S19; PARTIAL.  
 KW Ribosomal protein; Chloroplast; rRNA-binding.  
 FT NON TER 6  
 SQ SEQUENCE 6 AA; 732 MW; 63333735A411C000 CRC64;

Query Match 0.4%; Score 3; DB 10; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 1e+06;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 TRS 355  
 DB 1 TRS 3

RESULT 11  
 Q8KMS3 PRELIMINARY; PRT; 7 AA.  
 ID Q8KMS3;  
 AC Q8KMS3;  
 DT 01-OCT-2002 (TRENBLrel. 22, Created)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
 DE Putative MerX2 protein.  
 GN MERR2.  
 OS Klebsiella sp. LS13-39.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Klebsiella.  
 OX NCBI\_TaxID=143776;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LS13-39;  
 RX MEDLINE=21604134; PubMed=11763242;  
 RA Mindlin S.Z., Khlodii G.Y., Gorlenko Z.M., Minakhina S.V.,  
 RA Minakhin L.S., Kalyseva E.S., Kopteva A.V., Petrova M.A.,  
 RA Yurieva O.V., Nikiforov V.G.;  
 RT "Mercury resistance transposons of Gram-negative environmental  
 RT bacteria and their classification."  
 RL Res. Microbiol. 152:811-822(2001).  
 DR EMBL; AJ302776; CAC82975.1;  
 SQ SEQUENCE 7 AA; 608 MW; 6DC1B5BDD87DD6F0 CRC64;

Query Match 0.4%; Score 3; DB 2; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1e+06;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 AGA 103  
 DB 2 AGA 4

RESULT 12  
 Q47477 PRELIMINARY; PRT; 7 AA.  
 ID Q47477;  
 AC Q47477;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)  
 DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)  
 DE Tpi protein (Fragment).  
 GN TPI  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 [1]  
 RP SEQUENCE OF 7-7 FROM N.A.  
 RX MEDLINE=85203917; PubMed=3158524;  
 RA Hellings H.W., Evans P.R.;  
 RT "Nucleotide sequence and high-level expression of the major  
 RT Escherichia coli phosphofructokinase.";  
 RL Eur. J. Biochem. 149:363-373(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Evans P.;  
 RL Submitted (OCT-1986) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; X02519; CAA26359.1;  
 FT NON TER 1  
 SQ SEQUENCE 7 AA; 773 MW; 7416D33DDDB1DB0 CRC64;

Query Match 0.4%; Score 3; DB 2; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1e+06;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 EAA 231  
 DB 2 EAA 4

RESULT 13  
 P72081 PRELIMINARY; PRT; 7 AA.  
 ID P72081;  
 AC P72081;  
 DT 01-FEB-1997 (TRENBLrel. 02, Created)  
 DT 01-FEB-1997 (TRENBLrel. 02, Last sequence update)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
 DE 3'-methylcephem hydroxylase (Fragment).  
 GN CEFF.  
 OS Nocardia lactamdurans.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Pseudonocardiaceae; Pseudonocardiaceae; Amycolatopsis.  
 OX NCBI\_TaxID=1913;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=96009872; PubMed=7557411;  
 RA Coque J., Perez-Llarena F.J., Enguita F.J., Fuente J.L., Martin J.F.,  
 RA Liras P.;  
 RT "Characterization of the cmcH genes of Nocardia lactamdurans and  
 RT Streptomyces clavuligerus encoding a functional 3'-hydroxymethylcephem  
 RT O-carbamoyltransferase for cephamycin biosynthesis.";  
 RL Gene 162:21-27(1995).  
 DR EMBL; Z21682; CAA79797.1;  
 FT NON TER 1  
 SQ SEQUENCE 7 AA; 746 MW; 75A1B2CDD1E681C0 CRC64;

Query Match 0.4%; Score 3; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred.No.1e+06; 0; Indels 0; Gaps 0;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 663 AVT 665  
Db 4 AVT 6

## RESULT 14

Q8KMS9 PRELIMINARY; PRT; 7 AA.  
AC Q8KMS9;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Putative transposase (Fragment).  
GN TNIA.  
OS Enterobacter sp. CH2-4.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Enterobacter.  
OX NCBI\_TaxID=143777;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CH2-4;  
RX MEDLINE=21604134; PubMed=11763242;  
RA Mindlin S.Z., Kholodii G.Y., Gorlenko Z.M., Minakhina S.V.,  
RA Minakhin L.S., Kalyaeva E.S., Kopteva A.V., Petrova M.A.,  
RA Yurieva O.V., Nikiforov V.G.;  
RT "Mercury resistance transposons of Gram-negative environmental  
bacteria and their classification."  
RL Res. Microbiol. 152:811-822(2001).  
DR EMBL; AJ302778; CAC83058.1; -.  
FT NON TER 1  
SQ SEQUENCE 7 AA; 966 MW; 737B1B1046DAA9A0 CRC64;

Query Match 0.4%; Score 3; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred.No.1e+06; 0; Indels 0; Gaps 0;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 572 IEE 574  
Db 4 IEE 6

## RESULT 15

O07354 PRELIMINARY; PRT; 7 AA.  
AC O07354;  
DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE NifK (Fragment).  
GN NIFK.  
OS Synechococcus sp. (strain PCC 8801 / RF-1) (Cyanothecae PCC 8801).  
OC Bacteria; Cyanobacteria; Chroococcales; Cyanothecae.  
OX NCBI\_TaxID=41431;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RF-1;  
RX MEDLINE=99231861; PubMed=10217509;  
RA Huang T.C., Lin R.F., Chu M.K., Chen H.M.;  
RT "Organization and expression of nitrogen-fixation genes in the aerobic  
nitrogen-fixing unicellular cyanobacterium Synechococcus sp. strain  
RF-1."  
RL Microbiology 145:743-753(1999).  
DR EMBL; AF003700; AAC35193.1; -.  
FT NON TER 1  
SQ SEQUENCE 7 AA; 849 MW; 7412C72AA9D5B030 CRC64;

Query Match 0.4%; Score 3; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred.No.1e+06; 0; Indels 0; Gaps 0;  
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 692 DLV 694  
Db 4 DLV 6

Search completed: August 17, 2004, 10:16:38  
Job time : 45 secs

This Page Blank (uspto)